NEW HISTORY MATCHING METHODOLOGY FOR TWO PHASE RESERVOIR USING EXPECTATION-MAXIMIZATION (EM) ALGORITHM

A Dissertation

by

OLUWAFEMI OPEYEMI BALOGUN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Eduardo Gildin
Co-chair of Committee,	Ding Zhou
Committee Members,	Hadi Nasrabadi
	Samiran Sinha
Head of Department,	Daniel Hill

December 2017

Major Subject: Petroleum Engineering

ABSTRACT

The Expectation-Maximization (EM) Algorithm is a well-known method for estimating maximum likelihood and can be used to find missing numbers in an array. The EM Algorithm has been used extensively in Electrical and Electronics Engineering as well as in the Biometrics industries for image processing but very little use of the EM Algorithm has been seen in the Oil and Gas industry, especially for History Matching. History matching is a non-unique matching of oil rate, water rate, gas rate and bottom hole pressure data of a producing well (known as Producer) as well as the bottom hole pressure and liquid injection of an injecting well (known as Injector) by adjusting reservoir parameters such as permeability, porosity, Corey exponents, compressibility factor, and other pertinent reservoir parameters. EM Algorithm is a statistical method that guarantees convergence and is particularly useful when the likelihood function is a member of the exponential family. On the other hand, EM algorithm can be slow to converge, and may converge to a local optimum of the observed data log likelihood function, depending on the starting values.

In this research, our objective is to develop an algorithm that can be used to successfully match the historical production data given sparse field data. Our approach will be to update the permeability multiplier, thereby updating the permeability of each unobserved grid cell that contributes to the production at one or more producing wells. The EM algorithm will be utilized to optimize the permeability multiplier of each contributing unobserved grid cell.

DEDICATION

This work is dedicated to my wife, Kehinde Balogun (Mrs.), and my son, Eyinitan Balogun for their unflinching support and prayers for the many years spent working on my doctorate degree. Also thanking my mother, Grace Balogun (Mrs.) for a mother's prayer and support.

.

ACKNOWLEDGEMENTS

I also like to appreciate my committee members, Dr. Eduardo Gildin (Chair), Dr. Ding Zhou (Co-Chair), Dr. Hadi Nasrabadi and Dr. Samiran Sinha for their guidance and support throughout the course of this research project.

Thanks also go to my friends and colleagues and the department faculty and staff for making my time at Texas A&M University a great experience.

Finally, thanks to my wife and son for their encouragement, patience, prayers, support and love; and to my mother for her prayers and encouragement.

CONTRIBUTORS AND FUNDING SOURCES

This work was supervised by a dissertation committee consisting of Dr. Eduardo Gilding (Chair), Dr. Ding Zhou (Co-Chair), and Dr. Hadi Nasrabadi of the Department of Petroleum Engineering and Dr. Samiran Sinha of the Department of Statistics.

The Matlab software code that calculates the relative contribution of each grid cell to each producing well was provided by Feyisayo Olalotiti-Lawal of the Department of Petroleum Engineering.

All other work conducted for the dissertation was completed by the student independently.

There are no outside funding contributions to acknowledge related to the research and compilation of this document.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

Introduction

Reservoir engineers have a wide variety of methods of history matching, both gradient based and non-gradient based methods. However, these methods do not use the knowledge of the relative contributions of each grid cells towards a producer in the history matching process.

In this research, we assume that as little as 5% of the field permeability is known and are termed as observed grid cells. This will include, but not limited to, the permeability of grid cells that contain the wells. In other words, we assume sparse data scenario. The remaining 95% grid cell permeabilities are assumed to be zero and are termed as unobserved grid cell. We then make an initial guess of these 95% grid cells that was assumed unobserved using log-normal Kriging. We run our forward model to determine the oil and water production rate, the water-oil ratio, the mean error relative to the historical total production rate and the error relative to the historical water-oil ratio. The water-oil ratio error is then normalized and used along with the mean error of the total product rate to determine the initial cell-by-cell permeability multiplier for each contributing unobserved grid cell. Utilizing these permeability multipliers at this point to calculate new cell permeabilities with the process repeated until the mean production error is below some threshold value is what we have termed the Normalized Error method. However, the permeability multipliers are further run through the EM

Algorithm to get an optimum permeability multiplier for each contributing unobserved grid cell and in-turn used to calculate the cell permeability. This process is repeated until the mean production error is below some threshold value. This method of utilizing the EM Algorithm to optimize the outcome of the Normalized Error calculation is what we have termed the EM Algorithm method. The observed grid cells are assumed correct and not modified throughout the entire process.

Matching the total production and water-oil ratio in an oil-water system implies that given the two equations with two unknowns (oil rate and water rate), we can uniquely solve for them.

The results show our ability to determine a field-wide permeability distribution that provides a match to historical production on both the Egg model and the SPE 10 model, for the normalized error and normalized error optimized with EM Algorithm. We compare both methods to the Design of Experiment (DOE) method and show that both methods provide equally good history matches as the DOE method.

However, history matching has its inherent challenges, many of which will be highlighted as follows:

- Lack of reliable field data: We match historical fluid rate and pressure data using wellbore and near wellbore data without much knowledge of what the permeability and porosity distributions are for an entire field. In the current day and age of the "Factory Model" in which wells are drilled as quickly as possible without much coring or data collection, the issue of lack of field data has only been exacerbated. In many practical problems in the oil and gas industry, there are many cases in which core data from a well

is used as data for other wells many miles away. With the industry realizing the problem posed by the non-existence of field data, many companies launched efforts to try to make sense of sparse data, turning to Artificial Intelligence in the process. One such effort was undertaken by Yogesh et al [**84**] in which Artificial Neural Network (ANN) was used to calculated and infer missing data. However, ANN requires training by using actual data and/or by using related data with inferential relationship. Unfortunately, ANN doesn't perform well when the training data set is not representative of the entire scenario for which the ANN is trained to recognize. In other words, ANN doesn't perform well in scenarios with sparse data. This is seen by the 50% success rate of the history match of the well production as noted in the paper. Although in some cases, a good history match may be arrived at, in many cases, production forecast is usually off the mark. Having a more robust approach for deriving the missing data is very important, given the inevitable sparse data situation we are always face with.

- Non-uniqueness of the solution: History matching is inherently a non-unique inverse problem which in many cases results in good history matches but very poor production and pressure prediction **[105]**.
- Large computational runtime: A large computational runtime is seen when history matching large complex reservoirs with a considerably large number of uncertainty parameters that need to be adjusted to match historical data **[59, 64]**.

For this research, our objective is to automatically match historical production given sparse permeability data by developing an algorithm that can be used for

successful history matching with very limited or sparse data using the EM Algorithm to optimize grid cell permeability multipliers. We will use the Egg model **[85]** and the SPE10 **[86]** model as proxy for benchmarking the new proposed algorithm.

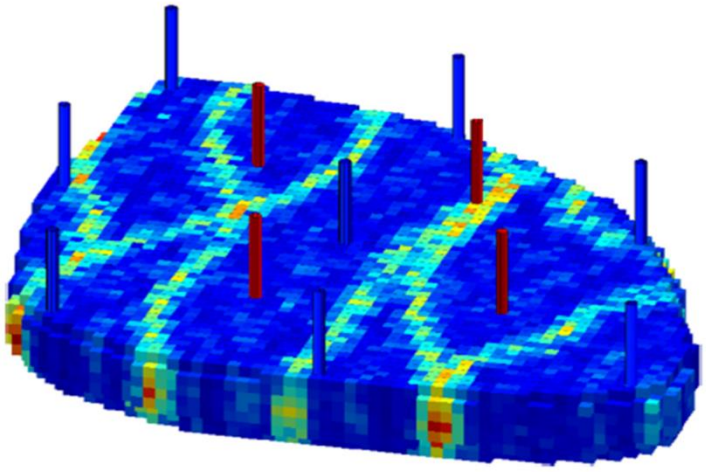


Figure 1: A realization of the Egg model (Reprinted from [85])

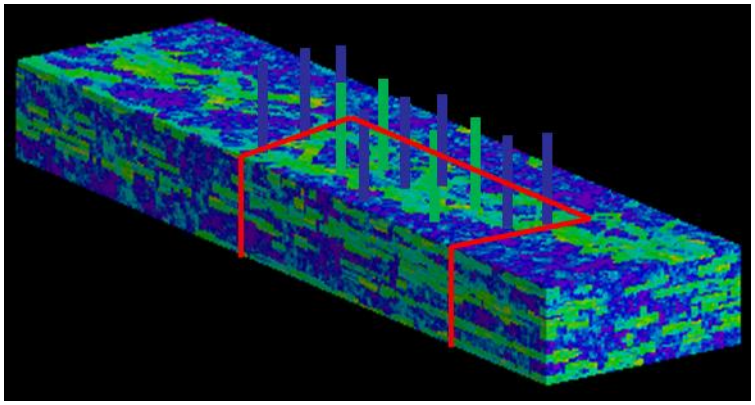


Figure 2: SPE10 Comparative Study showing the subsection and well locations for research study (Reprinted from [86])

Literature Review

This chapter is devoted to looking at the works that have been done using the EM Algorithm both in the Oil and Gas industry, however little, as well as in other industries to give an idea of its potential application to our particular case of history matching in oil reservoirs. Later chapters will be devoted to understanding the theoretical basis of The EM Algorithm, its mathematics and applications.

Nakayasu et al [**1]**, use The EM Algorithm to determine the value of permeability and porosity in a heterogeneous reservoir, an important work in the use of EM Algorithm for history matching. They utilized a combination of Value of Information, Gaussian random field models, Monte Carlo and the EM Algorithm for their work. Four petro-elastic facies (shale, organic-rich shale, limestone and sandstone) of the Marcellus Shale was classified by Schlanser et al [**2]** by applying Expectation Maximization to measure well logs. Uncertainty estimation of a pre-stack seismic inversion with L1 constraints was done by Sassen et al [**3]** in which they provided an improved earth model reconstruction over Gaussian or Tikhonov regulated inversions. Only these three works pertaining to the oil and gas industry were found to have used the EM Algorithm at the time this dissertation was written.

As mentioned earlier, the EM Algorithm has been used extensively in Electrical and Electronics Engineering as well as in the Biometrics industries for image processing. In the field of Electrical and Electronics Engineering, Einicke et al [**9]** discussed a navigation application in which the use of parameters estimated using the EM Algorithm improved filtering performance. They showed that the design and observed error

covariances are monotonically dependent on the residual error variance in the Expectation step while the residual error variance is monotonically dependent on the design and observed error covariances in the Maximization step. Hudson et al [**25**] in their paper used an Ordered Subsets EM, having an order of magnitude acceleration over the EM Algorithm, for image restoration from projections. Forero et al [**26]** developed a decentralized EM Algorithm to estimate the parameters of a mixture density model for use in distributed learning tasks performed with data collected at spatially deployed wireless sensors. They called this method the Consensus-Based Distributed EM Algorithm. EM type algorithm was used by Andrieu et al **[28]** to estimate static parameters in nonlinear non-Gaussian state space models which they found do not degenerate over time contrary to the standard Sequential Monte Carlo methods. On the other hand, Frenkel et al **[31]** investigated the application of the EM Algorithm to the estimation of time varying parameter estimation problem of Multiple Target Tracking (MTT) for a known number of targets. They noted that the EM algorithm achieve a linear dependency over conventional algorithms used to solve similar problems, problems that have computational complexity that depends exponentially on the number of targets. Lipinski et al **[29]** used the EM Algorithm to reconstruct Positron Emission Tomography using anatomical magnetic resonance information. They experimented with multiple EM algorithms and found that the Gauss-EM, which applies a Gauss function with the same mean for all pixels in a given anatomical region, though more sensitive to a priori information was more effective in noise reduction. Similarly, Ranganath et al **[50]** developed a multigrid EM algorithm to reconstruct positron

emission tomography. Along the same line, Levitan et al **[51]** utilized the EM algorithm to reconstruct emission tomography based on Poisson distribution of the statistically independent components of the image and measurement vectors, and extending their work to reconstructing a maximum a posteriori image using a multivariate Gaussian a priori probability distribution of the image vector. Likewise, the SPECT (Single-Photon Emission Computed Tomography) images where reconstructed using the EM Algorithm considering Compton scattering when calculating the photon detection probability matrix by Bowsher et al **[33]**. Image reconstruction was also done for diffuse optical tomography using sparsity regularization and EM Algorithm by Cao et al **[34]**. Another example of image reconstruction is one by Tom et al **[36]**, who reconstructed a high-resolution image by simultaneous registration, restoration and interpolation of low-resolution images using the EM Algorithm. Tom et al **[41]** further combined and extended their work to the simultaneous blur identification and restoration of multiple image planes that are obtained by an imaging system that measure the same scene using multiple sensors. The technique of detecting superimposed signals in code-division multiple access (CDMA) communication systems, which by itself poses a computation complexity which is exponential in the number of users, was developed by Barron et al **[42]** using the EM Algorithm. Legendijk at al **[43]** also investigated the identification and restoration of noisy blurred image without knowing the point-spread function of the degrading system, as well as the variance of the observed noise and the model of the original image, using the EM Algorithm. Carson et al **[46]** succeeded in retrieving

images from large and varying collections by utilizing the EM Algorithm using image content as a key – a very challenging problem.

Real-time Three-dimensional maps of indoor environments were reconstructed with a mobile robot using the EM Algorithm by Thrun et al **[30]**. This work showed that EM can be used to reconstruct missing data in real-time. Marinakis et al **[37]** were able to infer sensor position and qualitative topological map of an environment given a set of cameras and with non-overlapping fields of view without prior knowledge of the environment nor the exact positon of sensors within the environment to understand traffic patterns. Georghiades **[32]** combined the EM Algorithm with the Viterbi Algorithm was used to detect unsynchronized symbol sequence when the symbol timing information was missing.

The absorption energy distribution of molecular probes was calculated from their adsorption isotherms without prior knowledge of the distribution function, isotherm, or isotherm data smoothening by Stanley et al **[39]** using the EM Algorithm. They demonstrated a high stability and convergent towards the maximum-likelihood estimates of the absorption energy distribution. Sergeev A.S. et al [**24]** showed that the EM Algorithm can be used to estimate the maximum likelihood of gametic frequencies of multilocus polymorphic codominant systems based on sampled population data, a feat difficult to achieve because the estimation of the frequencies of multilocus gametes based on the data of multilocus genotypes is sometimes impossible due to their incompleteness of the data. Excoffier et al **[48]** utilized the EM Algorithm to identify the gametic phase of haplotypes which is usually unknown when diploid individuals are

heterogeneous at more than one locus. The evaluated the performance of the EM Algorithm for simulated data representing both DNA sequences and highly polymorphic loci with different levels of recombination and found that the EM Algorithm performed best for large samples regardless of recombination rates among loci. Bailey et al **[47]** extended the EM Algorithm to identify motifs in unaligned biopolymer sequences by using sub-sequences which actually occur in the biopolymer sequences as starting points for the EM Algorithm to increase the probability of finding the globally optimal motifs. Similar work was done by Lawrence et al **[52]** in which the EM Algorithm was used to identify and characterize binding sites in a set of unaligned DNA fragments, also known as biopolymer sequences.

This review of the work done using the EM Algorithm, particularly for missing data resolution and image inversion problems both in the biotech and electrical electronics industries show that extensive work has been done using the EM Algorithm. But very little of such work has been done in the oil and gas industry as noted earlier in this chapter, particularly as it applies to reconstructing the permeability distribution of an oil reservoir given the oil and water production of the producing wells.

It is indeed a possibility to reconstruct field-wide permeability distribution from an initial guess of permeability distribution using the EM Algorithm to arrive at a relatively good match between the historical production data and calculated production data derived with the permeability distribution reconstructed using EM Algorithm.

Research Objective and Scope

As earlier stated, our objective is to automatically match historical production given sparse permeability data by developing an algorithm that can be used for successful history matching with very limited or sparse data using the EM Algorithm to optimize contributing grid cell permeability multipliers. We achieve this by:

- utilizing log-normal Kriging to make an initial guess of what the unknown permeabilities are for the unknown grid cells
- developing new analytical equations to calculate the average permeability multiplier of each contributing grid cell that take into consideration both the distance of the grid cells from the producers as well as the relative contribution of each grid cell to each producer, which takes into consideration the geology of the earth model.
- Applying the EM Algorithm, we optimize these average permeability multipliers and calculate new field-wide permeabilities.

For this research, we will only focus on matching historical oil and water production rates in an incompressible oil-water system with bottom-hole pressure constraint by modifying only the permeability of the contributing grid cells, on a cell-by-cell basis, using the optimum permeability multiplier derived with the EM Algorithm.

Thesis Organization

This dissertation is organized as follows: In chapter 1, we introduce the objective of this research and an overview of how we have solved the stated objected. We undertake a review of what has been done, both in the oil and gas industry as well as other industries, with EM algorithm. We stated how we intend to use the EM Algorithm in this research and stated the scope of this research.

We review history matching as well as a brief overview of some of the history matching methods used in the oil and gas industry as well as a few of the sources of uncertainties experienced in history matching in chapter 2.

In chapter 3, we share a brief overview of the EM Algorithm, its pros and cons, and how we have applied it in this research. We show a step-by-step approach of our new algorithms and workflows that show both the Normalized Error method and the EM Algorithm method for calculating the permeability multipliers of contributing grid cells.

Both the Normalized Error method and the EM Algorithm method were used to history match the Egg Model **[85]** and the SPE10 model **[86]** and the compared the history match achieved to the history match achieved with Design of Experiment in chapter 4. We also highlighted some of the identified restrictions of the new algorithm.

We conclude in chapter 5 by giving a brief overview of what we have achieved in this research. We also highlight future research opportunities for the new proposed history matching algorithms.

CHAPTER II

REVIEW OF HISTORY MATCHING

History Matching

History matching is a non-unique inverse problem of adjusting a set of parameters, which in themselves have a measure of inherent uncertainty, to match the historical production of a well and the entire field so as to forecast future production. The objective is to minimize the difference between the observed and numerically calculated data according to a measure of the "goodness" of the match. As earlier mentioned and as noted by Yamada T. **[74]** and Cancelliere M. et al **[64],** this is a process that gives non-unique results, given the fact that multiple combinations of the parameters of a reservoir will yield similar production trend.

History matching can be a time consuming and computationally intensive effort with a wide range of uncertainty which is not without its own issues as noted by Tavassoll Z. et al **[66]**. There have been many attempts to reduce the time and computational needs as well as the uncertainty it entails to narrow down on the model matches as quickly as possible **[92, 93, 94]**.

The general algorithm for the inversion process is as stated by Dadashpour M. **[60]** and illustrated in Figure 3.

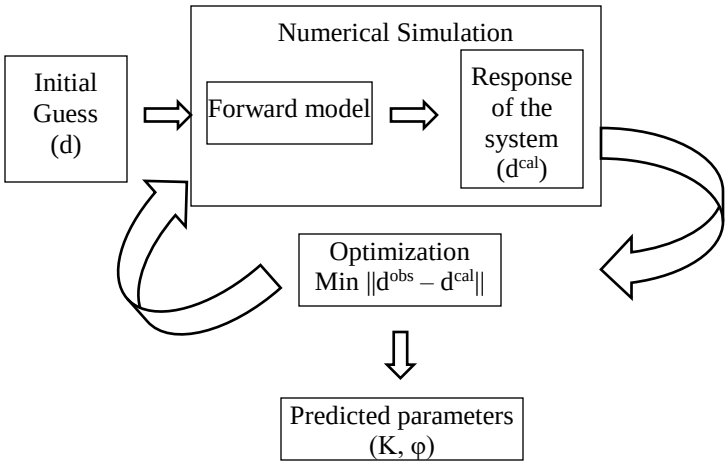


Figure 3: Common algorithm for the inversion process (Reprinted from [60])

From Figure 3 above, the numerical simulation is the running of a forward model to get a system response. Here, a system of discretized and linearized nonlinear partial differential equations with the appropriate boundary conditions that represent the physics of the fluid flow in a reservoir is solved with the aid of a computer.

Given that our main objective is to determine unknown parameters of an oil reservoir, we show here the system of partial differential equations used in the numerical simulation of multi-phase flow in Equation 2.1 through Equation 2.12. The reader may refer to classical books in numerical simulation of flow through porous media **[95, 96]** For a black oil model multi-phase flow of fluids in a reservoir, the flow equation in Cartesian coordinates is given as:

$$\frac{\partial}{\partial x}\left[\beta_c \frac{A_x k_x k_{ro}}{\mu_o B_o}\left(\frac{\partial p_o}{\partial x} - \gamma_o \frac{\partial Z}{\partial x}\right)\right]\Delta x + \frac{\partial}{\partial y}\left[\beta_c \frac{A_y k_y k_{ro}}{\mu_o B_o}\left(\frac{\partial p_o}{\partial y} - \gamma_o \frac{\partial Z}{\partial y}\right)\right]\Delta y + \frac{\partial}{\partial z}\left[\beta_c \frac{A_z k_z k_{ro}}{\mu_o B_o}\left(\frac{\partial p_o}{\partial z} - \gamma_o \frac{\partial Z}{\partial z}\right)\right]\Delta z = \frac{V_b}{a_c}\frac{\partial}{\partial t}\left(\frac{\emptyset S_o}{B_o}\right) - q_{osc} \quad \ldots\ldots\ldots\ldots\ldots \quad 2.1$$

$$\frac{\partial}{\partial x}\left[\beta_c \frac{A_x k_x k_{rw}}{\mu_w B_w}\left(\frac{\partial p_w}{\partial x} - \gamma_w \frac{\partial Z}{\partial x}\right)\right]\Delta x + \frac{\partial}{\partial y}\left[\beta_c \frac{A_y k_y k_{rw}}{\mu_w B_w}\left(\frac{\partial p_w}{\partial y} - \gamma_w \frac{\partial Z}{\partial y}\right)\right]\Delta y + \frac{\partial}{\partial z}\left[\beta_c \frac{A_z k_z k_{rw}}{\mu_w B_w}\left(\frac{\partial p_w}{\partial z} - \gamma_w \frac{\partial Z}{\partial z}\right)\right]\Delta z = \frac{V_b}{a_c}\frac{\partial}{\partial t}\left(\frac{\emptyset S_w}{B_w}\right) - q_{wsc} \quad \ldots\ldots\ldots\ldots\ldots \quad 2.2$$

$$\frac{\partial}{\partial x}\left[\beta_c \frac{A_x k_x k_{rg}}{\mu_g B_g}\left(\frac{\partial p_g}{\partial x} - \gamma_g \frac{\partial Z}{\partial x}\right) + \beta_c R_s \frac{A_x k_x k_{ro}}{\mu_o B_o}\left(\frac{\partial p_o}{\partial x} - \gamma_o \frac{\partial Z}{\partial x}\right)\right]\Delta x + \frac{\partial}{\partial y}\left[\beta_c \frac{A_y k_y k_{rw}}{\mu_w B_w}\left(\frac{\partial p_w}{\partial y} - \gamma_w \frac{\partial Z}{\partial y}\right) + \beta_c R_s \frac{A_y k_y k_{ro}}{\mu_o B_o}\left(\frac{\partial p_o}{\partial y} - \gamma_o \frac{\partial Z}{\partial y}\right)\right]\Delta y + \frac{\partial}{\partial z}\left[\beta_c \frac{A_z k_z k_{rw}}{\mu_w B_w}\left(\frac{\partial p_w}{\partial z} - \gamma_w \frac{\partial Z}{\partial z}\right) + \beta_c R_s \frac{A_z k_z k_{ro}}{\mu_o B_o}\left(\frac{\partial p_o}{\partial z} - \gamma_o \frac{\partial Z}{\partial z}\right)\right]\Delta z = \frac{V_b}{a_c}\frac{\partial}{\partial t}\left(\frac{\emptyset S_g}{B_g} + \frac{\emptyset R_s S_o}{B_o}\right) - q_{gsc} \quad \ldots\ldots\ldots\ldots\ldots \quad 2.3$$

With

$$q_{gsc} = q_{fgsc} + R_s q_{osc} \quad \ldots\ldots\ldots\ldots\ldots \quad 2.4$$

Where for $i = o, w, g$ and $\eta = x, y, z$

P_i = Oil/Water/Gas Pressure

S_i = Oil/Water/Gas Saturation

γ_i = Oil/Water/Gas Specific Gravity

B_i = Oil/Water/Gas Formation Volume Factor

μ_i = Oil/Water/Gas Viscosity

q_{isc} = Oil/Water/Gas Production Rate at Standard Conditions

q_{fgsc} = Free Gas Production Rate at Standard Conditions

k_{ri} = Relative Permeability to Oil/Water/Gas

k_η = Absolute Permeability normal to x/y/z direction

A_η = Cross-Sectional Area normal to x/y/z direction

R_s = Solution Gas-Oil Ratio

$\emptyset$ = Porosity

V_b = Bulk Volume

Z = Elevation referred to datum

β_c = Transmissibility Conversion Factor

a_c = Volume Conversion Factor

Since porosity, $\emptyset$ is constant, for an incompressible fluid,

$$\frac{\partial}{\partial t}\left(\frac{\emptyset S_i}{B_i}\right) = 0 \qquad \ldots\ldots\ldots\ldots\ldots \qquad 2.5$$

for $i = o, w, g$

For scenario in which gravity is assumed negligible,

$$\frac{\partial Z}{\partial j} = 0 \qquad \ldots\ldots\ldots\ldots\ldots \qquad 2.6$$

where $j = x, y, z$

And

$$\sum_{i=1}^{3} S_i = 1 \quad \text{for } i = o, w, g \qquad \ldots\ldots\ldots\ldots\ldots \qquad 2.7$$

$$P_{cow} = P_o - P_w = f(S_w) \qquad \ldots\ldots\ldots\ldots\ldots \qquad 2.8$$

$$P_{cgo} = P_g - P_o = f(S_g) \qquad \ldots\ldots\ldots\ldots\ldots \qquad 2.9$$

Assume that a reservoir is sliced up and represented by discrete three dimensional grid blocks in the *i*, *j* and *k* directions where *i* = 1, 2, ….., n_x; *j* = 1, 2, ….., n_y; and *k* = 1, 2, ….., n_z.

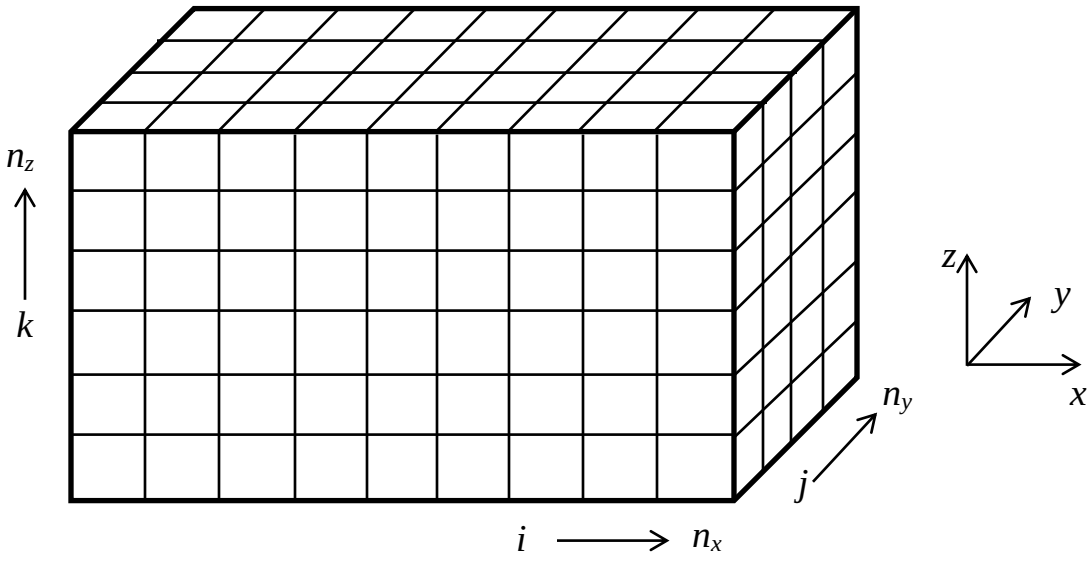


Figure 4: Reservoir split into grid blocks for simulation

The multi-phase partial differential flow equations 2.1 to 2.3 are discretized to represent the flow in each grid block and are solved applying equations 2.7 to 2.9 to determine the oil, water and gas rates as the fluid flows through the grid cells.

The system of partial differential equations may be solved using any of the following schemes - Finite Difference, Finite Element and Finite Volume **[97, 98, 99, 100]**.

The entire process of history matching is an attempt to find out what the field permeability distribution k_x, k_y, and k_z are by plugging into equation 2.1 through 2.3 subject to initial and boundary conditions, guesses for k_x, k_y, and k_z, and other unknown variables, solving for pressure, and the rates q_o, q_w, and q_g at each timestep for the entire length of time of the historical data subject to either the bottom hole pressure constraint or rate constraint. At the end of the simulation, the calculated pressure and rates are compared with the measured pressure and rate data to determine the deviation of the calculated data from the historical data – this is known as the error. If the error is larger than some error threshold, further guess for k_x, k_y, and k_z (and possibly other parameters) is made and the model is again run for the length of time of the historical data to re-calculate the error and determine where it is larger or smaller than the error threshold limit. This is done repeatedly until the calculated error is lower than the error threshold limit. Then a history match is said to be achieved. Sometimes, a correlation between permeability and porosity is used to back calculate the porosity for a given guess of permeability. The usual practice is to simultaneously make a guess at porosity and permeability when attempting to match the historical data.

As might be imagined, conducting a history matching manually can be a tedious and tricky process and can sometimes be seen as art than science. This is the reason why many methods of history matching have been developed. Many such methods include, but not limited to, Streamlines **[81]**, Genetic Algorithm (GA) **[58, 59]**, Ensemble Kalman Filter (EnKF) **[101]**, Bayesian methods **[102]**, and Assisted History Matching (AHM) using Design Of Experiment (DOE) **[103]**. For this dissertation, the methods of history matching with Streamlines and Assisted History Matching using DOE will be explained.

The general method of history matching starts with having a base model and making changes to parameters within the model to arrive at matches or near-matches of the historical production and pressure data.

Although the purpose of this research is not to analyze the different methods of history matching, it is worth mentioning the different methods that are currently being used. Readers who are interested in further exploring any of these methods have a vast library of materials to read from, some of which are Shahkarami A. et al **[57,58]**, Hajizadeh Y. et al **[59,61]**, Aissaoui K. **[62]**, Yang P.H. et al **[63]**, Sayyafzadeh M. **[68]**, Singh A. et al **[69]**, Almeida Netto S.L. et al **[70]**, Abdel-Rahman M.R. et al **[73]**, MacMillan D.J. et al **[75]**, Gavalas G.R. et al **[76]**, Tzu-hao Y. et al **[77]**, Cheng **[81]**. This is by no means an exhaustive literature search and should be seen as such. The history matching methods currently used are broadly categorized into the following:

- Manual history matching: This is mainly based on a trial and error approach of manually turning the knobs of many uncertainty parameters such as porosity and

particularly the permeability using permeability multipliers, with the hope of matching historical fluid rate and pressure data. This, depending on the complexity and size of the reservoir as well as the number of wells, can be a frustrating ordeal and also time consuming, sometimes taking several months of work.

- Assisted/Automatic history matching: Today, many reservoir simulation engineers utilize Assisted History Matching (AHM) which is a combination of Automatic history matching and human intervention. In general, the degree of human intervention is the difference between Assisted History Matching and Automatic History Matching and for many practitioners, there is not a clear distinction between the two. Here the algorithm simply tries to minimize a misfit function to obtain the model that best approximates the historical production and pressure data. It has its drawbacks as noted by Cancelliere M. et al **[64]** and Kabir C.S. et al **[67]**. Examples of this history matching approach are AHM with Design of Experiment and Streamline simulation.

Assisted History Matching (AHM) Using Design of Experiment

This method gives the ability to make multiple guesses of the input parameters, running the resulting model to calculate the production rates and pressures with the hope of getting at least one match. With this method, the following steps are taken:

1. We determine the variables in equation 2.1 to 2.3 that we intend to adjust in order to match the historical data as the uncertain parameters, ζ_1 through ζ_n.
2. The range (Maximum and Minimum) of uncertainties for each parameter ζ_i is determined for all i = 1 to n
3. The range of uncertainties for each parameter ζ_i may be distributed within a certain distribution if assumed to be a continuous variable e.g. Gaussian, Triangular, or Uniform; or not fitted with a distribution but treated as a discrete variable.
4. If the number of trial runs is chosen as *m*, which is the number of guesses they wish to make, and a parameter has a range of uncertainties distributed according to a given distribution, the m guesses of each parameter will be fitted accordingly within that distribution. Therefore, for the uncertain parameter β_i, there will be m different guess, all distributed according to some defined distribution within the range of uncertainties for each parameter ζ_i.
5. This leaves us with an "m x n" matrix as shown below

$$\begin{bmatrix} \zeta_{1,1} & \cdots & \zeta_{1,n} \\ \vdots & \ddots & \vdots \\ \zeta_{m,1} & \cdots & \zeta_{m,n} \end{bmatrix}$$

Each guess, *m*, is a model with *n* uncertain parameters. If at the end of *m* runs, no history match is found, changes may be made to the range of uncertainties of one or more parameters, or changes made to the distribution used to fit the *m* guesses for each parameter. This process is repeated until a history match is achieved. This method, to

some extent, is a trial and error process that hopes that at least one of the *m* combination of *n* unknown parameters will bring about a good match with the historical data. Although Design of Experiment is easy to set up, practitioners will need to understand the nature of each variable, ζ_i, so as to determine the appropriate distribution to use. Also, practitioners will need to consult with Subject Matter Experts to determine the range of uncertainty of each variable, ζ_i. Not taking time to do these might result in situations in which the practitioner might get a match with completely erroneous inputs.

Automatic History Matching (AHM) Using Streamlines

According to Cheng **[81,104]**, automatic history matching using streamlines uses streamline-derived sensitivities to update geologic models. He further describes the four major steps of automatic history matching using streamlines as follows:

1. Run an initial streamline-based flow simulation to compute production response at the wells.
2. Quantify the mismatch between observed and computed production response.
3. Run a streamline-based analytic sensitivity computation of the production response with respect to reservoir parameters, ζ_i for $i = 1$ to n.
4. Update the reservoir properties to match the production history via inverse modeling using streamline-derived sensitivities.

These four major steps are further outlined in his workflow shown in Figure 5 below.

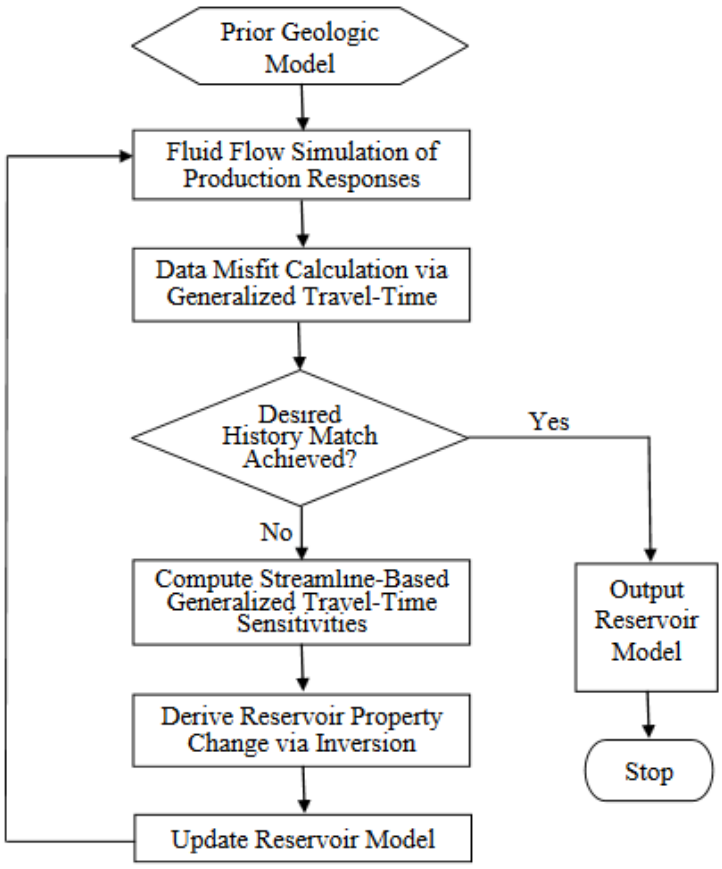


Figure 5: Flowchart for automatic history matching with streamline (Reprinted from [81])

Streamline method only attempts to match historical data by matching breakthrough (arrival) time of the displacing fluid by generalized travel time inversion and not by matching the actual historical production data set. It works best when history matching is done with total liquid rate as constraint. Streamline method is particularly

useful for waterfloods for matching water breakthrough and the fractional flow of water and utilizes the injected water as a tracer.

Uncertainty and History Matching

Uncertainty is an inevitable issue we have to live with in history matching and adds more complexity to the process of history matching. This stems from the fact that there are an unknown number of processes and interactions going on several thousands of feet below the earth surface that we are unable to determine with 100% certainty how these processes and interactions affect well production as well as pressure distribution across a reservoir.

There are many sources of uncertainty in history matching. These can be broadly categorized into four categories:

- Geology: This is the greatest source of uncertainty and is impossible to be eliminated given the fact that geologic interpretation via seismic and microseismic interpretation cannot be done with absolute certainty makes this issue quite significant. Although considerable efforts have been made over the past several years to improve seismic and microseismic data gathering and interpretation, we still have plenty of work to do in this area given that we can only infer from images, sometimes at low resolution, what might be in existence thousands of feet below the earth's surface.
- Data quality and quantity: This is another major source of uncertainty. The manner in which data is acquired and interpreted, especially PVT and core data;

the type of mud that was used in drilling the well; the frequency of well tests performed on the wells to ascertain production coming from the wells; the accuracy of production back allocation in a commingled scenario; etc. These and many more affect the quantity and quality of data we have available for analysis and consequently, the level of uncertainty we have to deal with during history matching.

- Scale-up: Much of the data utilized in our simulation and history matching are derived at the microscopic and used to determine data at macroscopic and gigascopic scales. As an example, porosity and permeability are measured from tiny cores with lengths and diameters in inches taken from a few well locations. These microscopic scale porosity and permeability are used as an average porosity and permeability for an entire reservoir which is at the macroscopic and gigascopic scale of several thousands of feet. It is clear that these core derived porosity and permeability do not represent the entire reservoir, yet these are the numbers that are input into the simulation, introducing a significant amount of uncertainty into our simulation and history matching.
- Mathematical: This is generally the least uncertainty we have to deal with in history matching as this is a result of the rounding errors of the discretization and linearization of the mathematical equations with which the numerical simulator calculates production rates and pressure. In this research, we have used the method of finite volume as described by Shahvali M. et all **[80]** to determine the relative contributions of each grid cell towards production at each of the

producers. One major drawback of this finite volume method is that it introduces error due to diffusion and therefore the calculated relative contributions, which will be discussed later in this work, inherently have some measure of uncertainty in them.

As a result of these uncertainties in history matching, it is generally advised to give ranges of possibility from a distribution by stipulating the P10, P50 and P90 scenarios for production rate and pressure forecasts derived from history match. This is seen as a safety net underscoring the uncertainty of the data.

Much work has been done in the area of integrating uncertainty during history matching. Schiozer D.J. et al **[71]** proposed a 7-step methodology for integrating uncertainty during history match which they recommend for fields with complex behavior and when uncertainty remain significant even after acquiring production data.

Opportunities

Given the uncertainty challenges previously mentioned as pertain to history matching, the following opportunities arise:

- Having a robust analytical technique to calculate the permeability distribution for an entire field is very valuable especially when developing a field management plan or an asset development plan to determine where to drill and land a well.

- Having an automatic history matching method that will optimize the unknown parameters (permeability) towards achieving a minimum measure of production error.
- Ability to reconstruct an optimum field-wide permeability distribution given very sparse permeability data for few locations in the field.
- Although the non-uniqueness of history matches cannot be eliminated, effort will be made to find methodology that reduces the amount of uncertainty and non-uniqueness by matching both historical production rates and the WOR in an oil-water system.

In the following chapter, given an initial sparse permeability data, we show a detailed step-by-step approach of how we have developed and used a new analytical method to calculate and optimize the field-wide permeability distribution for a field using the EM Algorithm.

CHAPTER III

EXPECTATION MAXIMIZATION AND OUR NEW APPROACH

The Expectation Maximization (EM) Algorithm

The Expectation Maximization (EM) Algorithm was explained and given its name in a 1977 paper by Arthur Dempster, Nan Laird, and Donald Rubin **[4]**. Prior to the 1977 paper, a very detailed look into the EM method for exponential families was published by Rolf Sundberg in his thesis and in technical papers **[6]**. The Dempster-Laird-Rubin paper **[4]** in 1977 generalized the method and demonstrated a convergence analysis for a wider class of problems. The convergence analysis of the Dempster-Laird-Rubin paper was flawed and a correct convergence analysis was published by C.F. Jeff Wu **[14]** in 1983 who established the EM method's convergence outside of the exponential family.

The EM Algorithm is an iterative method for finding maximum likelihood or maximum a posteriori (MAP) estimates where the model depends on unobserved latent variables. The method alternates between an Expectation (E) step, which creates a function for the expectation of the log-likelihood evaluated using the current estimate for the parameters; and a Maximization (M) step, which maximizes the expected log-likelihood derived in the E-step. The EM Algorithm is particularly good for resolving Gaussian Mixtures. Interested readers may refer to Chuong et al. **[82]** to learn more about the coin toss example and how the EM Algorithm is used to solve the coin toss problem. In the study, a depiction of the popular coin toss example calculation using the

EM Algorithm is illustrated in an iterative process to arrive at the probability of a coin being either in a Coin A bucket or a Coin B bucket. It shows the capability of the EM Algorithm in determining which coin belongs to which bucket without any prior knowledge of the buckets the coins belong to.

Theoretical Basis of EM Algorithm

In this dissertation, we will not cover the mathematical formulations of the EM Algorithm in-depth. We will however, provide enough information to introduce the EM Algorithm as is used in the formulation of the proposed algorithm. The EM Algorithm procedure, as utilized in this research, is described as follows:

Assume observed data *X* - In our case, given that we will be optimizing the permeability multiplier, *X* would be the permeability multiplier matrix, λ, which will be a row vector of all *L* permeability multipliers. In its general form,

$$X = \begin{bmatrix} x_{1,1} & \cdots & x_{1,L} \\ \cdot & \cdot & \cdot \\ x_{\upsilon,1} & \cdots & x_{\upsilon,L} \end{bmatrix} = (\hat{x}_1, \ldots, \hat{x}_L) \qquad \text{---------- 3.1}$$

Where

υ = number of parameters; *L* = number of data; and each $x_{i,j}$ is a member of vector $\hat{x}_j$.

In our case, υ would be 1 since we are only optimizing permeability multiplier vector, λ. In a situation where we have multiple variables to optimize, υ would be the number of variables to be optimized. For a 5 by 5 grid block, L would be 25.

For any given multivariate Gaussian mixture, the probability distribution function (pdf) is given as

$$\Gamma_d(x) = \sum_{k=1}^{d} \frac{\xi_k}{\sqrt{(2\pi)^s|\Sigma_k|}} exp\left[-\frac{1}{2}(x-\mu_k)^T\Sigma_k(x-\mu_k)\right] = \sum_{k=1}^{d} \xi_k N(x|\mu_k, \Sigma_k) \quad \text{---------- 3.2}$$

Where

x = Permeability multiplier

d = number of Gaussian distributions

μ_k = υ-dimensional mean vector

Σ_k = υ x υ covariance matrix of the kth Gaussian

Such that

$$\xi_{1,} \dots\dots, \xi_d \geq 0 \text{ and } \sum_{k=1}^{d} \xi_k = 1 \quad \text{---------- 3.3}$$

which are the positivity condition and the normalization condition respectively.

z_{ik} is the probability that a given permeability multiplier x_i (λ) of an unobserved grid cell belongs to some cluster k.

For the E-step, we first calculate the probability that x_i belongs to some cluster, k

$$z_{i,k} = \frac{\xi_k N(x_i|\mu_k, \Sigma_k)}{\sum_{j=1}^{s} \xi_j N(x_i|\mu_j, \Sigma_j)} \quad \text{---------- 3.4}$$

hence, we are able to calculate the expectation of the log-likelihood by first guessing an initial probability $z_{i,k}$, denoted as $z_{i,k}^{(0)}$.

$$Q_d(X|\mu,\Sigma,\xi) = \sum_{i=1}^{L}\sum_{k=1}^{d} z_{i,k}^{(0)} ln(\xi_k N(x_i|\mu_k,\Sigma_k)) \quad \text{---------- 3.5}$$

The M-step maximizes the log-likelihood, Q over $\mu_k, \Sigma_k$and ξ_k to solve

$$\frac{\partial Q}{\mu_k} = 0 \; ; \quad \frac{\partial Q}{\Sigma_k} = 0 \; ; \quad \frac{\partial Q}{\xi_k} = 0$$

Solving these results in an iteration to calculate the μ_k, $\Sigma_k$and ξ_k

$$\mu_k^{(t)} = \frac{\sum_{i=1}^{L} z_{i,k}^{(t-1)} x_i}{\sum_{i=1}^{L} z_{i,k}^{(t-1)}} \quad \text{---------- 3.6}$$

$$\Sigma_k^{(t)} = \frac{\sum_{i=1}^{L} z_{i,k}^{(t-1)} \left(x_i - \mu_k^{(t)}\right)\left(x_i - \mu_k^{(t)}\right)^T}{\sum_{i=1}^{L} z_{i,k}^{(t-1)}} \quad \text{---------- 3.7}$$

$$\xi_k^{(t-1)} = \frac{1}{L}\sum_{i=1}^{L} z_{i,k}^{(t-1)} \quad \text{---------- 3.8}$$

In equation 3.5, we use the BIC to calculate the number of Gaussian distributions, *d*. The objective is to pick the *d* that makes BIC a minimum using the following:

$$BIC = -2lnQ_d + \alpha lnL \quad \text{---------- 3.9}$$

Where α is the number of parameters in the probability distribution function Γ_d above.

The EM Algorithm – Pros and Cons

The EM Algorithm tends to be very good for solving complex non-linear inverse problems and resolving incomplete data. It can be used in cases where data values are missing and does not require computation of gradient/Hessian of likelihood function unlike conjugate gradient, gradient descent or Gauss-Newton method. It is rather simple to implement and tends to be much more numerically stable than gradient based methods.

Pros

- Guaranteed to converge given that it searches within a bounded domain.
- Suitable for resolving incomplete data without a-priori knowledge of the data.
- Particularly useful when function is an exponential family.
- There is no need for large computational resources to compute any Hessian.

Cons

- EM algorithm can be slow to converge.
- May converge to a local maximum of the observed data likelihood function, depending on starting values.
- It can be arbitrarily poor in high dimensions and there can be an exponential number of local optima.

For this research, we use the EM Algorithm to optimize the permeability multiplier of grid cells so as to modify the permeabilities of those grid cells in order to match historical production from producing wells in an oil-water system. The details of the approach are explained further in the following sections.

The New Approach

To discuss the approach we have undertaken, we begin by showing the true permeability distribution of one of the layers of the Egg model in Figure 6 and explaining the step-by-step approach we have undertaken thereafter.

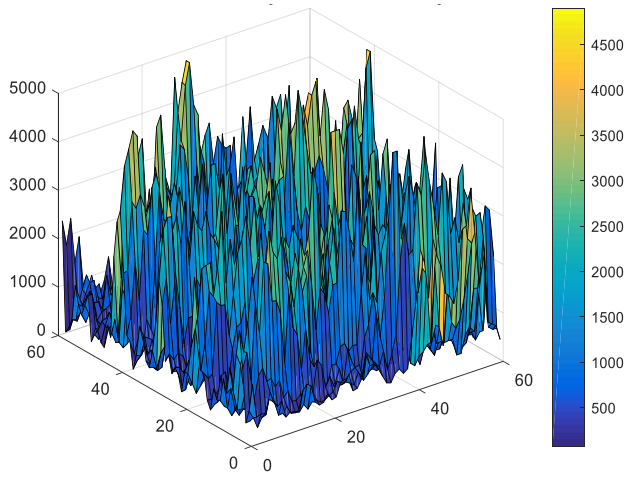


Figure 6: Permeability Distribution of a Layer in the Egg Model

In the approach, it is assumed that we only know 5% of the permeability of grid cells. This implies that 95% of the total number of grid cells would be considered unobserved. These grid cells that are assumed unobserved will henceforth be referred to as unobserved grid cells. Also, amongst these unobserved grid cells, there will be some grid cells that contribute to the production at a producer or a number of producers. These unobserved grid cells will be referred to as contributing unobserved grid cells. Figure 7 below depicts the initial permeability distribution for a layer of the Egg model with 95% missing data. Initially, all unobserved grid cells permeabilities are assigned an initial value of zero.

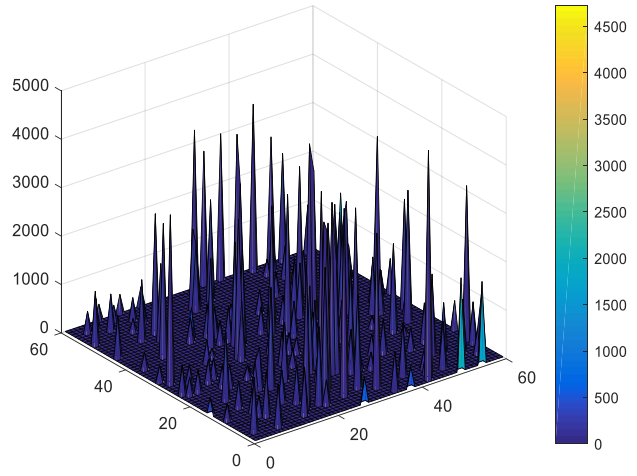


Figure 7: Initial Permeability Distribution with 95% of cell permeability unknown for same layer of the Egg model as in Figure 6

This distribution is then fed through log-normal Kriging to calculate an initial guess for the unobserved grid cells. It is important to note that the initial guess using

log-normal Kriging will be equivalent to an initial field wide permeability distribution one would receive from geologist who built the earth model. Also, it is important to note that the underlying assumption of ordinary Kriging data transformation is that the data set is normally distributed. During the course of this research, in addition to the log-normal Kriging approach, we attempted to generate the missing numbers using random number generation and the ordinary Kriging. It should be no surprise that both the random number generation and ordinary Kriging methods did not provide very good outputs.

For the random number generation, no spatial relationship between a sample point (observed grid cell) and unsampled point (unobserved grid cell) is taken into consideration. However, geologists would agree that earth models in general show stratigraphic trends within a formation except in cases where we have faults. In general, the notion that a data point is completely stratigraphically different and completely random from the next would be highly erroneously. Hence, it was no surprise that the use of the random approach did not yield permeability distributions that made much sense. Figure 8 below shows the result of generating an initial field wide permeability distribution using the random number generation approach.

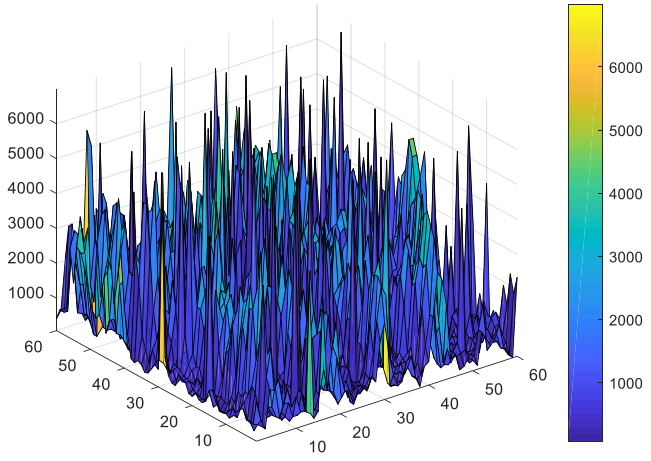


Figure 8: Permeability distribution estimates generated randomly for same layer of the Egg model as in Figure 6

For the ordinary Kriging, a linear Kriging methodology which takes into account the spatial relationship between a sample point (observed grid cell) and unsampled point (unobserved grid cell), it became clear that utilizing a method that is based on the normal distribution of data on permeability data set that is generally known to be log-normally distributed would also result in erroneous data. Also, given the highly variable nature of our data set, ordinary Kriging will be unable to providing better estimation of the spatial relationship between a sample point (observed grid cell) and unsampled point (unobserved grid cell). Figure 9 below shows the result of generating an initial field wide permeability distribution using ordinary Kriging.

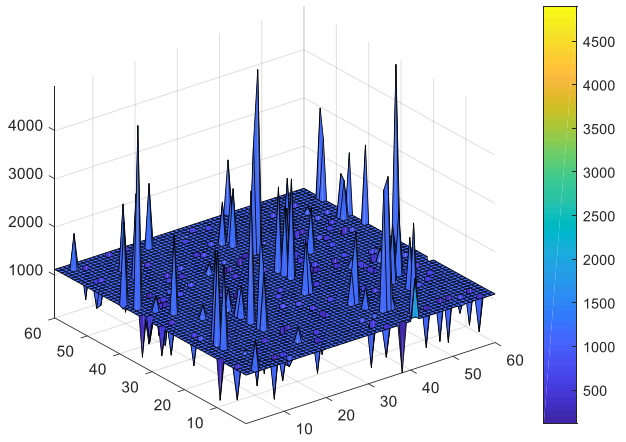


Figure 9: Permeability distribution estimates from ordinary Kriging for same layer of the Egg model as in Figure 6

On the other hand, given that permeability is log normally distributed, we applied log-normal Kriging to our initial distribution shown in Figure 7 above. In log-normal kriging, which is a non-linear Kriging methodology, we start by applying log transform to the initial data set before applying ordinary kriging to find the unsampled data (unobserved grid cells) and then applying an inverse log transform to the resulting data set. Log-normal Kriging is preferred to the ordinary Kriging because **[90]**:

(a) it reduces the variability between two points (the sampled and the unsampled data points) by the application of log transforms, thus providing better estimation of the spatial relationship between the two points

(b) the application of log transforms makes the log normally distributed data set to be normally distributed. By so doing, we will therefore only need to calculate the error variance to be able to describe the uncertainty of estimation of the unsampled data points.

Figure 10 below shows the result of the application of log-normal Kriging to the permeability distribution.

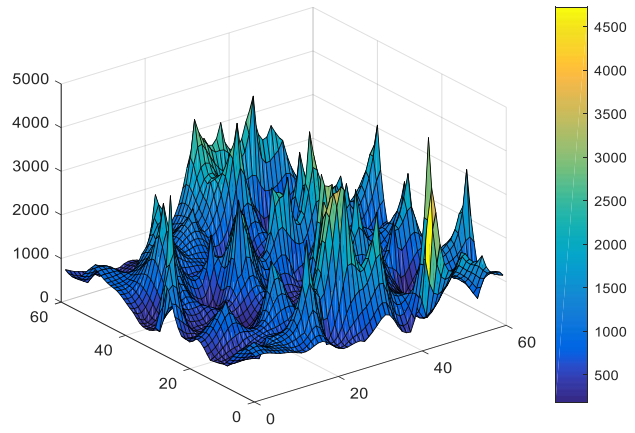


Figure 10: Permeability distribution estimates from log-normal Kriging for same layer of the Egg model as in Figure 6

The resulting permeability distribution from the log-normal Kriging is used to run a forward model to calculate the total production rates (in an oil-water system, the total production rate is the sum of the oil and water production rates) and water-oil ratios of producing wells. These total production rates and water-oil ratios are then fed through loops of error normalization and EM Algorithm to calculate cell-by-cell permeability multipliers to update the field wide permeability distribution and match the historical total production rate, q_T, and water-oil ratio, W, from which the oil rate and water rate can be back calculated as:

$$q_o = \frac{q_T}{1+W} \qquad \text{---------- 3.10}$$

And

$$q_w = q_T * \left(\frac{W}{1+W}\right) \qquad \text{---------- 3.11}$$

The resulting permeability distribution that gave the match with historical production is shown in Figure 11 below for one of the layers.

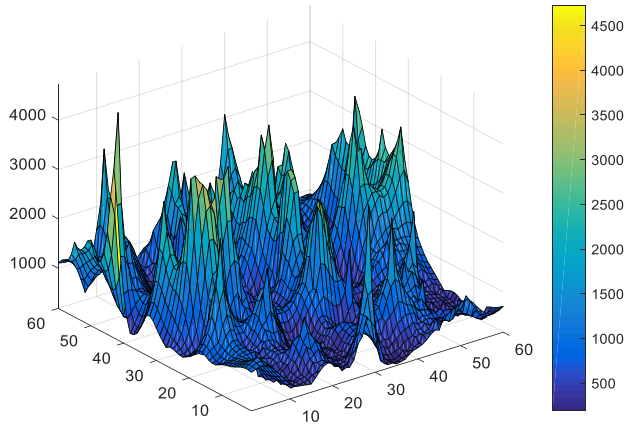


Figure 11: Permeability distribution for EM Algorithm matched history for same layer of the Egg model as in Figure 6

The approach taken in this research is based on the assumption that if the relative contributions of each cell to each producer is known, we should be able to back-calculate and adjust the cell permeabilities accordingly so as to match the production seen at the

producers. We use the finite volume technique proposed by Shahvali M. et al **[80]** to calculate the relative contribution of each cell to each producer.

It is important to note that this approach is non-linear and is complicated by the following:

1. A given cell may have contributions from more than one injector. As depicted in Figure 12 below, each grid cell contributes some percentage (ratio) of the production observed at a producer. These are hypothetically indicated as the numbers in the contributing grid cells in Figure 12 below. The cells bordered with the red lines have contributions from two injectors while those bordered with the green lines have contributions from all three injectors. To resolve this scenario and get the relative contribution of each cell, regardless of the number of injectors that flow into the cell, we utilize the method described by Shahvali M. et al **[80]** in which finite volume method was used as a post processing method to obtain flow diagnostics of the relative contributions of each contributing grid cell.
2. For a given grid cell, contribution to each producer is not equal and depends on the permeability of the fluid path between the cell and the producer; and the distance of fluid travel from the cell to the producers; amongst other possible dependencies.
3. For a given loop, described in the workflow below, depending on the flow direction and the permeabilities of the adjacent cells, the amount of fluid that flows out of a particular cell to the adjacent cells vary as depicted by the relative

thickness of the arrows in Figure 12 below. It is important to note that the relative contributions do not remain constant from one loop to the next given that the permeability of the contributing cells are altered during each loop run.

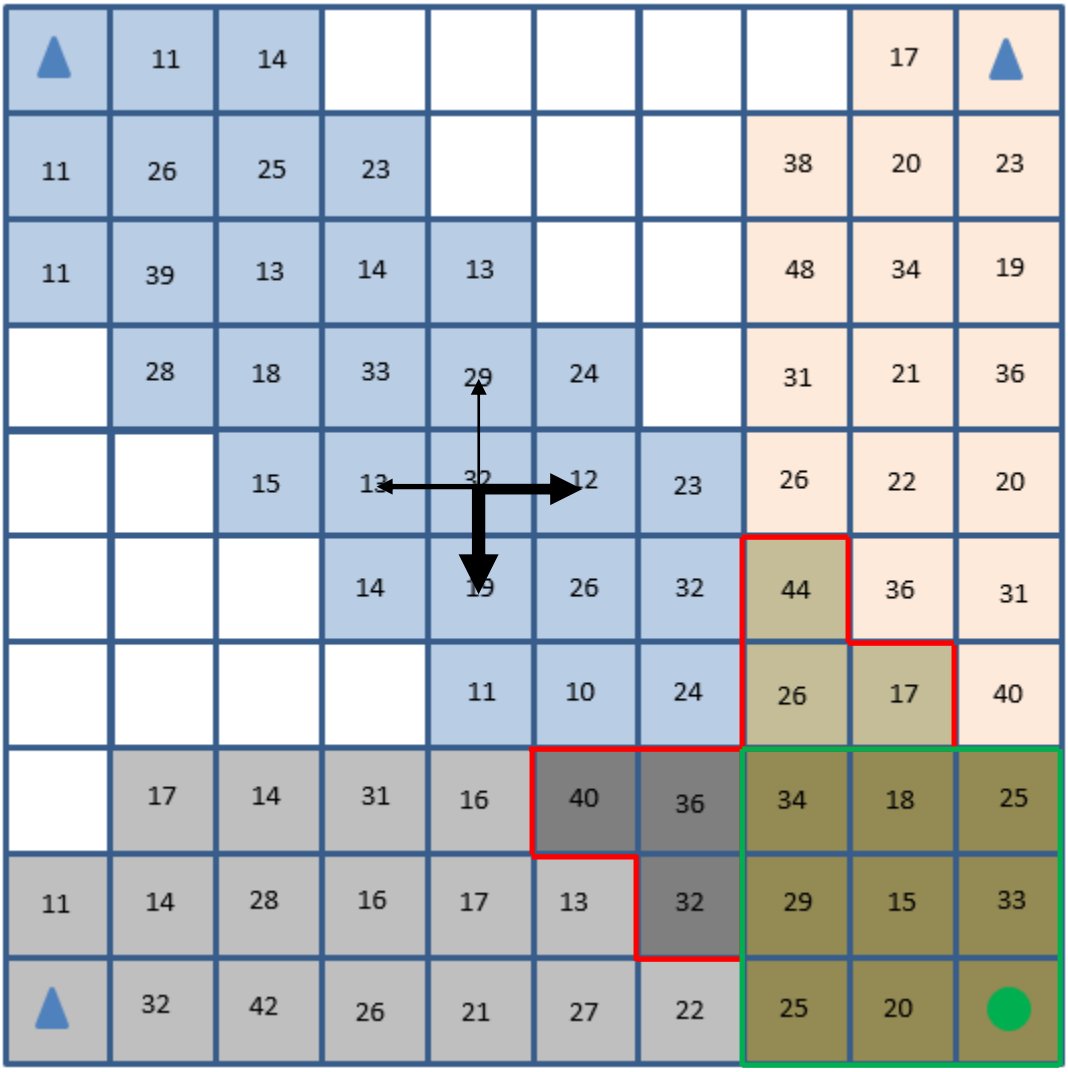


Figure 12: Depiction of relative contributions of each grid cell to a producer

The steps below show the new history matching method for both the Normalized Error method and the EM Algorithm method:

1. Calculate the historical total production as the sum of the historical oil and water production

$$\tilde{q}_{Ti} = \tilde{q}_{oi} + \tilde{q}_{wi} \qquad \text{---------- 3.12}$$

and historical Water-Oil ratio,

$$\widetilde{W}_i = \tilde{q}_{oi}/\tilde{q}_{wi} \qquad \text{---------- 3.13}$$

for each producer.

2. Get an initial guess for unobserved grid cell permeability using log-normal Kriging to develop an initial guess of the field-wide permeability distribution.
3. Run the resulting permeability distribution in the simulator to get a calculated total production as the sum of the oil and water production

$$q_{Ti} = q_{oi} + q_{wi} \qquad \text{---------- 3.14}$$

and Water-Oil Ratio

$$W_i = q_{oi}/q_{wi} \qquad \text{---------- 3.15}$$

for each producer *i*.

4. Calculate individual contributions of each grid cell to each producer, C_{ji} utilizing the method of Shahvali M. et al **[80]**. This also tells us which cell, *j*, affects which wells, *i*; j = 1 to n_v where $n_v = n_x * n_y * n_z$ and n_x, n_y, n_z, are number of grid cells in the x, y and z directions respectively.
5. Compare the calculate total production to the historical total production by calculating the mean error

$$\bar{E}_i = mean[(\tilde{q}_{Ti} - q_{Ti})/\tilde{q}_{Ti}] \text{ ---------- } 3.16$$

   a. If mean error, $\bar{E}_i > 0$, we are producing less than we ought to, hence production needs to be increased by increasing permeability of cells that affect that particular producer *i*.
   b. If mean error, $\bar{E}_i < 0$, we are producing more than we ought to, hence production needs to be decreased by decreasing permeability of cells that affect that particular producer *i*.
6. Calculate the normalized Water-Oil ratio error by calculating Water-Oil ratio error, normalized by the maximum Water-Oil ratio error

$$\bar{\bar{E}}_{Wi} = (\widetilde{W}_i - W_i)/\max(\widetilde{W}_i - W_i) \text{ ---------- } 3.17$$

7. For each contributing unobserved grid cell, calculate Lambda, λ, for each producer. E.g. for each active cell in a given field with 4 producers, we calculate 4 Lambdas, 1 for each producer as follows

$$\lambda_i = 1 + \text{sign}(\bar{E}_i) * \bar{\bar{E}}_{Wi} \quad \text{---------- 3.18}$$

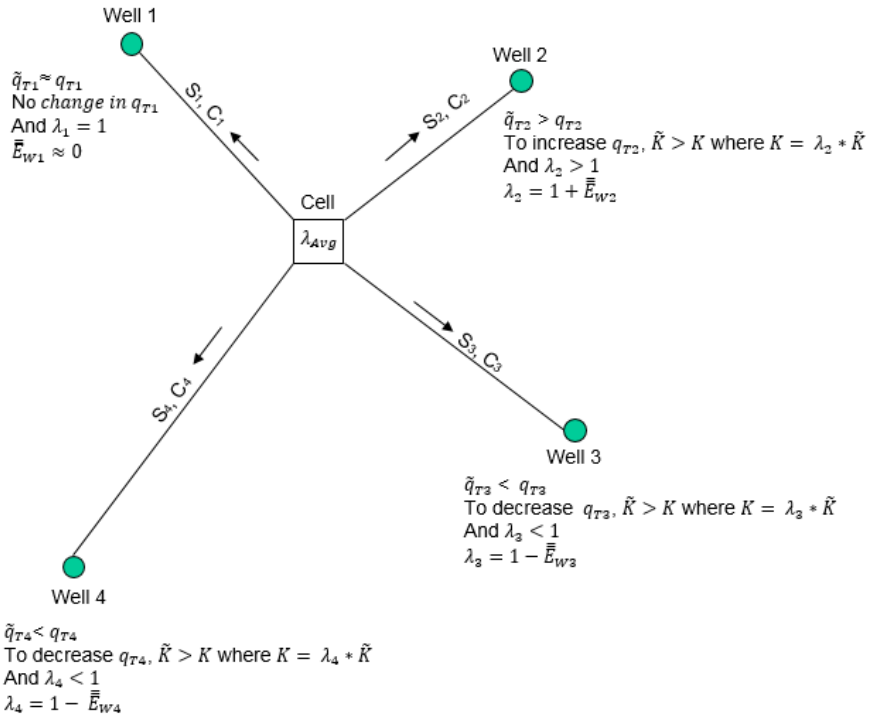


Figure 13: Calculating λAvg for a grid cell that contributes to production in four producers

This means that for each contributing unobserved grid cell, λ is weighted relative to a 1 (100%). If the calculated total production rate is more than the historical total production rate, the calculated total production rate will need to be reduced,

the sign($\overline{E}_i$) is negative so as to make λ less than 1 by the amount of error $\overline{\overline{E}}_{Wi}$. On the other hand, if the calculated total production rate is less than the historical total production rate, the calculated total product rate will need to be increased, the sign($\overline{E}_i$) is positive so as to make λ greater than 1 by the amount of error $\overline{\overline{E}}_{Wi}$. However, if the calculated total production rate is approximately the same as the historical total production rate, λ remains as 1 given that the error $\overline{\overline{E}}_{Wi} \approx 0$.

8. For each contributing unobserved grid cell, calculate an initial average Lambda as a function of both the distance of the unobserved grid cell from each producer, S_{ji}, and the individual contributions of each contributing unobserved grid cell to each producer, C_{ji}. Hence, for a given contributing unobserved grid cell, we calculate an initial average permeability multiplier

$$\lambda_{Avg} = \Sigma(\lambda_i S_{ji} C_{ji}) / \Sigma(S_{ji} C_{ji}) \quad \text{----------} \quad 3.19$$

This assures us that the initial average permeability multiplier calculated is based only on wells that receive contributions from the grid cell being that $C_{ji} = 0$ if there is no contribution to that producer *i* from the unobserved grid cell, *j*.
For the Normalized Error method, the initial average permeability multiplier, λ_{Avg}, now becomes our Permeability multiplier and calculate our new grid cell permeability.

$$K_{j,new} = \lambda_{Avg} * K_{j,prev} \quad \text{----------} \quad 3.20$$

At this point, we go back to step #3 above to run the simulator with the new permeabilities to determine whether error < δ. If error > δ, we continue the steps #3 through #8 until error < δ where δ is the error threshold limit.

However, to optimize the initial average permeability multiplier, we utilize the EM Algorithm method by proceeding with the following steps below.

9. To optimize the initial average permeability multiplier, we utilize the EM Algorithm and determine the upper and lower bounds as follows. For a given unobserved grid cell, n',

    a. If $\lambda_{Avg} < \lambda_{Prev}$, then

        i. $\lambda_{LBound} = \lambda_{Avg}$ ---------- 3.21a

        ii. $\lambda_{UBound} = \lambda_{Prev}$ ---------- 3.21b

    Therefore, our permeability multiplier bounds become $(\lambda_{Avg} , \lambda_{Prev}]$

    b. If $\lambda_{Avg} > \lambda_{Prev}$, then

        i. $\lambda_{LBound} = \lambda_{Prev}$ ---------- 3.22a

        ii. $\lambda_{UBound} = \lambda_{Avg}$ ---------- 3.22b

    Therefore, our permeability multiplier bounds become $[\lambda_{Prev} , \lambda_{Avg})$

    Where λ_{Prev} is the previous permeability multiplier

10. With the bounds for each grid cell, we randomly generate permeability multipliers (in our case, we chose to have 100 data points). These random permeability multipliers of the given grid cell are normally distributed and are projected on a 2D plane by plotting the random permeability multipliers vs. normalized permeability multipliers (normalized from 0 to 1) as shown below.

Figure 14 below shows the projection of a permeability multiplier bound (0.748,1] projected in 2D.

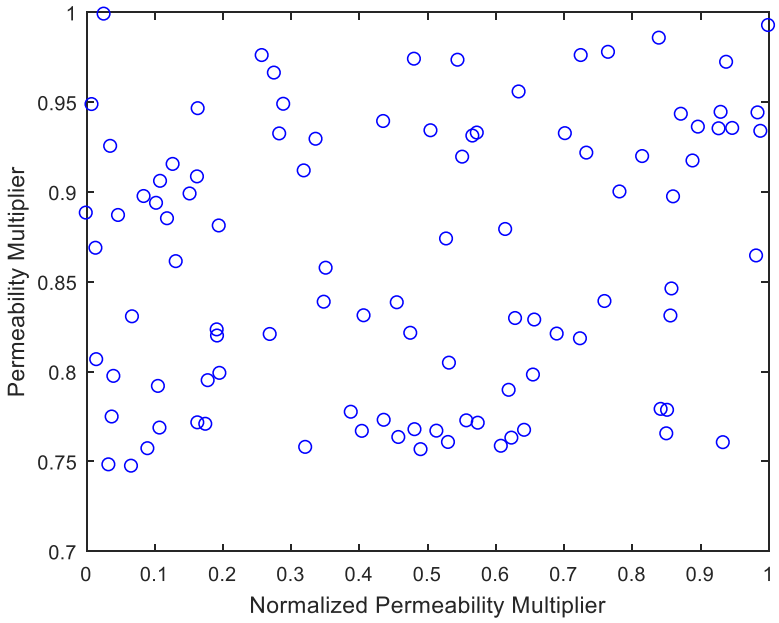


Figure 14: Random permeability multipliers projected in 2D plane

11. The EM algorithm partitions the random permeability multiplier data set into however many clusters, calculates the centroid of each cluster $\dot{k}_{\iota}$.

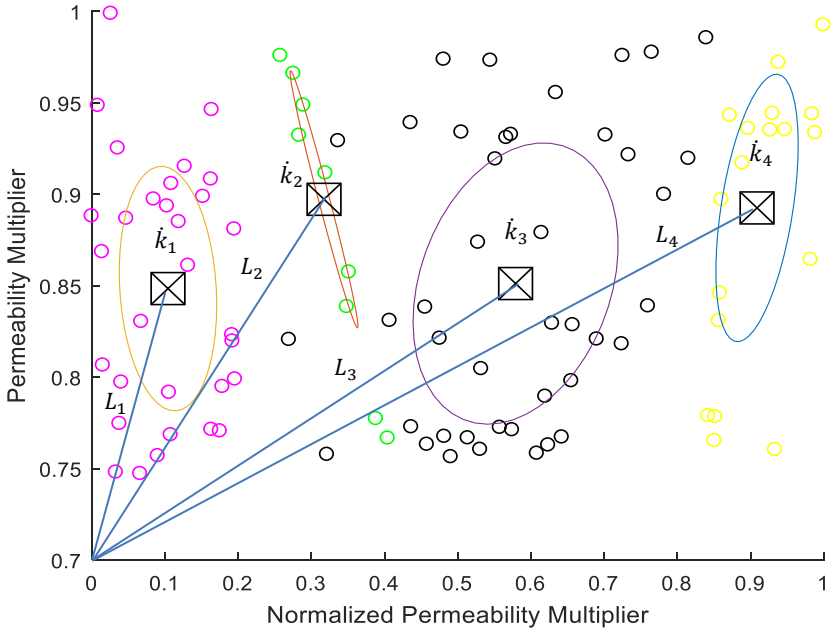


Figure 15: Centroids of the clusters represent cluster averages

12. The geometric mean of the centroids relative to the origin is calculated to determine the average permeability multiplier of the unobserved grid cell n'.

$$\bar{\lambda}_j = \frac{\sum_{l=1}^{n_c} \dot{k}_l L_l}{\sum_{l=1}^{n_c} L_l} \quad \text{---------- 3.23}$$

Where, in this case, n_c is the number of clusters and L_i is the distance from a common point in the cell to each of the centroids.

13. We step through all unobserved grid cells and calculate new permeability multiplier $\bar{\lambda}_j$ for each using the described EM algorithm approach.

14. We calculate the new permeabilities for each unobserved grid cell

$$K_{j,new} = \bar{\lambda}_j * K_{j,prev} \quad \text{---------- 3.24}$$

Where $K_{j,prev}$ is the previous permeability of grid cell j.

15. We go back to step #3 above to run the simulator with the new permeabilities to determine whether error < δ. If error > δ, we continue the steps #3 through #15 until error < δ where δ is the error threshold.

The EM algorithm process is summarized and portrayed on a 5 x 5 grid of Figure 16 in which after the resulting permeability multiplier bounds for a given cell are optimized in two clusters – Cluster 1 and Cluster 2. User may use more than 2 clusters but we've chosen to describe the method with only two clusters in Figure 16 below. The centroid of each cluster is taken as the permeability multiplier of each cluster as shown in step #11 above. A geometric mean of these cluster permeability multipliers relative to the cell origin is calculated as shown in step #12 above. The resulting mean permeability multiplier is taken as the permeability multiplier for that particular contributing grid cell, *j*.

Once again, for the Normalized Error method, steps #9 through #12 are omitted, and λ_{Avg} is used as the permeability multiplier instead of $\bar{\lambda}_j$.

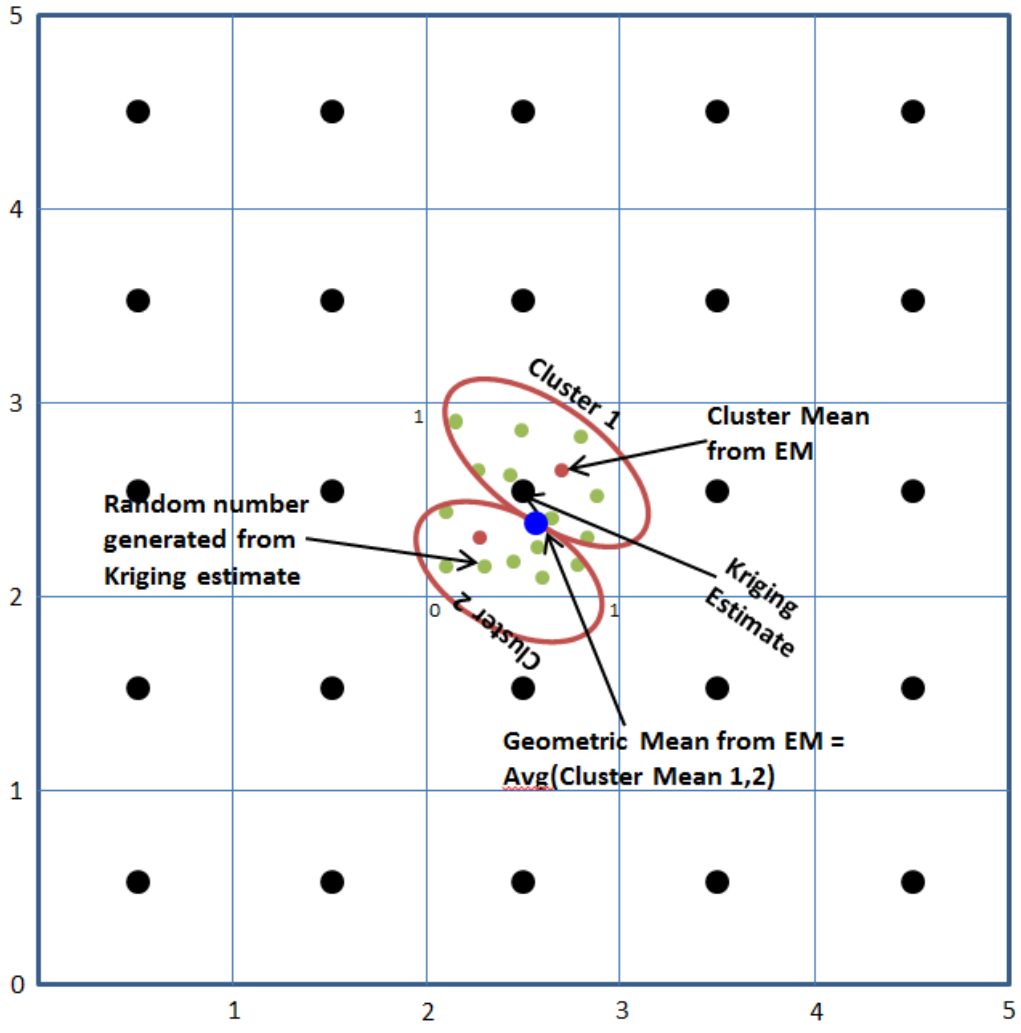


Figure 16: Summary depiction of the EM Algorithm to calculate permeability

Normalized Error History Matching Workflow

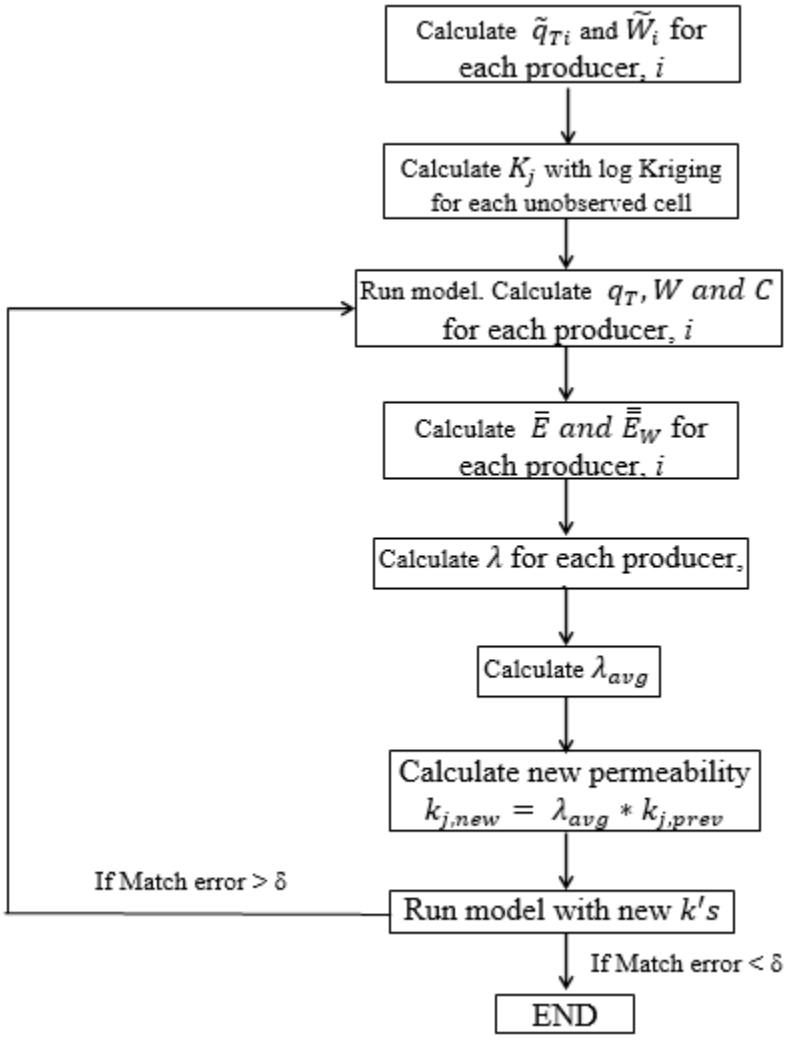


Figure 17: Normalized Error History matching workflow

EM Algorithm History Matching Workflow

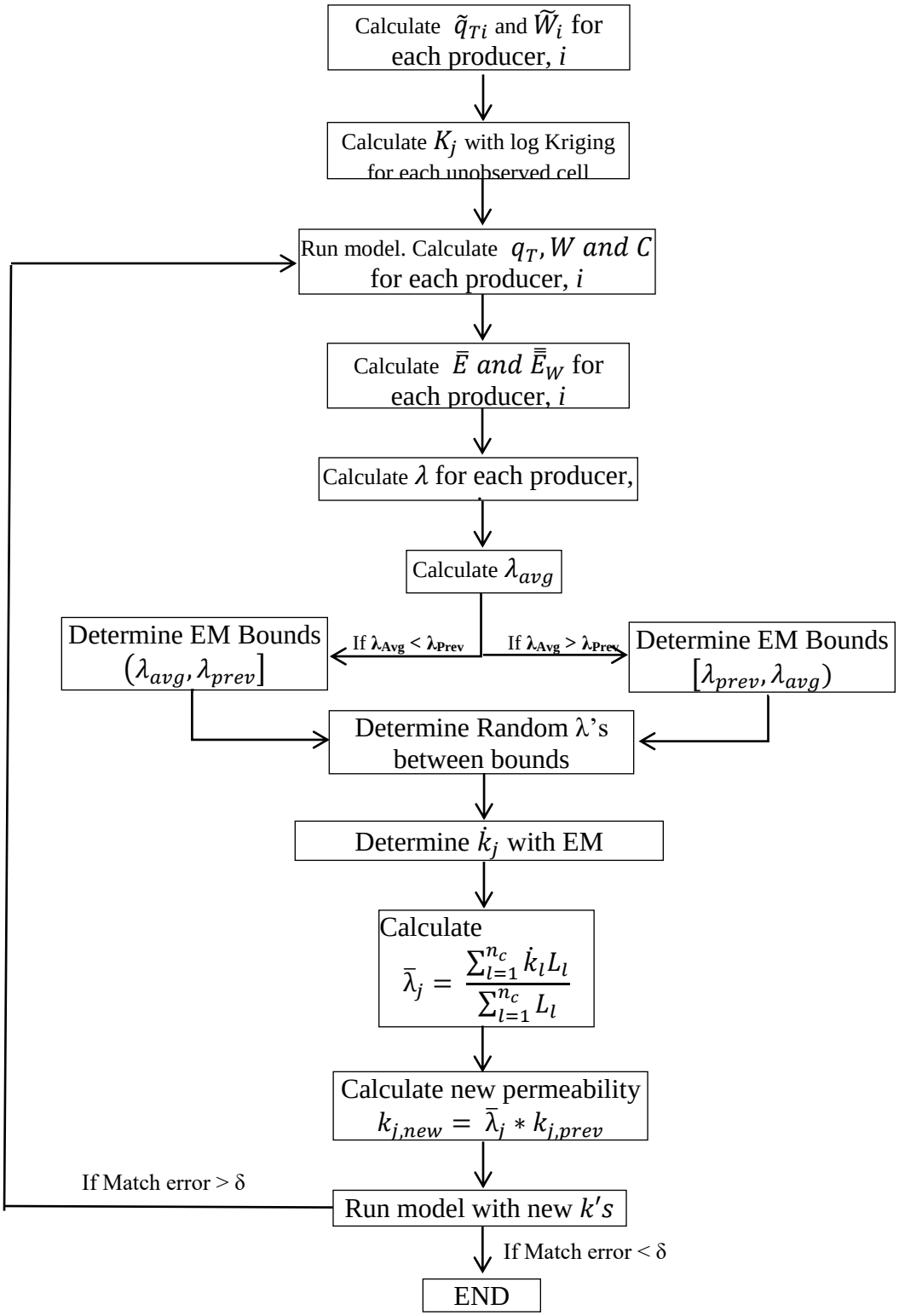


Figure 18: EM Algorithm History matching workflow

To further explain how the EM algorithm has been applied in this research, we start from step #9 above.

- Assume we have independent distributions of clusters of permeability multipliers indicated by the purple, green and cyan:

$$P_{dk}(x) = \frac{\xi_k}{\sqrt{(2\pi)^s|\Sigma_k|}} exp\left[-\frac{1}{2}(x-\mu_k)^T\Sigma_k(x-\mu_k)\right] = \xi_k N(x|\mu_k, \Sigma_k) \quad \text{---------- 3.25}$$

Where

x = Permeability Multiplier

d = number of Gaussian distributions

μ_k = *s*-dimensional mean vector of the kth Gaussian

Σ_k = *s* x *s* covariance matrix of the kth Gaussian

ξ_k = weight of the kth Gaussian

- Then, the joint distribution of the mixture, in red, is given as

$$\Gamma_d(x) = \sum_{k=1}^{d} \frac{\xi_k}{\sqrt{(2\pi)^s|\Sigma_k|}} exp\left[-\frac{1}{2}(x-\mu_k)^T\Sigma_k(x-\mu_k)\right] = \sum_{k=1}^{d} \xi_k N(x|\mu_k, \Sigma_k) \quad \text{---------- 3.26}$$

Such that

$$\xi_{1,} \ldots\ldots, \xi_d \geq 0 \text{ and } \sum_{k=1}^{d} \xi_k = 1 \quad \text{---------- 3.27}$$

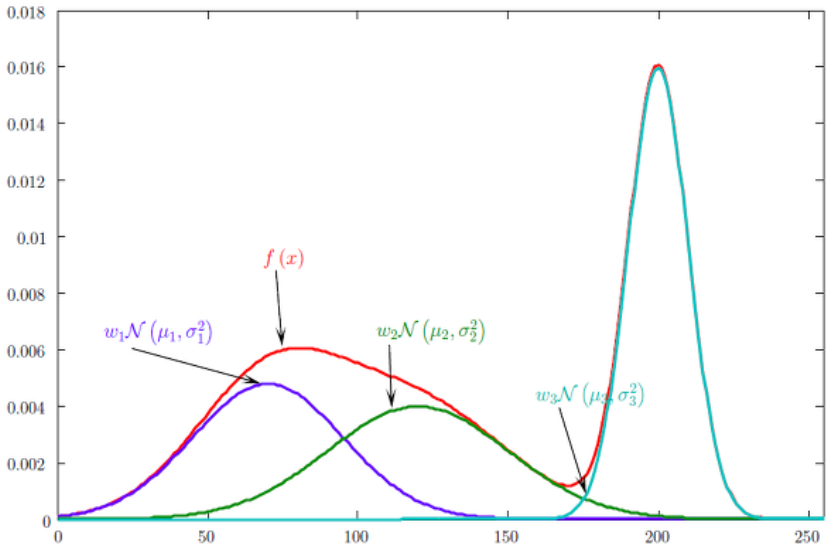


Figure 19: Example of a pdf of a Gaussian Mixture (Reprinted from [89])

- Now, assume we have the mixture of permeability multipliers and wish to bin them in their separate distributions (clusters) without prior knowledge of which cluster each belongs to. Hence, for any given multivariate Gaussian mixture, the probability distribution function (PDF) is given by the equation 3.25 above
- In other words, if we take a contributing unobserved grid cell as depicted in Figure 16, with the 2D projection within bounded domains (λ_{avg},λ_{Prev}] or [λ_{Prev}, λ_{avg}) as already described, we will have a grid cell with grid cells as seen in Figure 20 below.

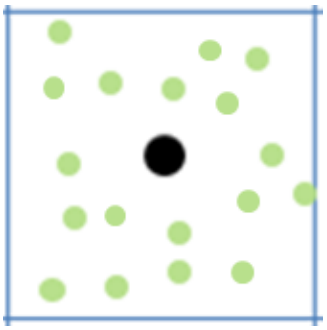


Figure 20: Grid cell depicting distribution of permeability multipliers within bounded domain

The objective is to find which cluster each data point (permeability multiplier) belongs to

- In the E-Step, we calculate the Expectation of the log-likelihood that a data point (permeability multiplier) x_i belongs to a cluster *k.* The likelihood that x_i belongs to distribution (cluster), *k*, is given as

$$\xi_k N(x_i|\mu_k, \Sigma_k) \quad \text{---------- 3.28}$$

Hence, the responsibility, z_{ik}, that x_i belongs to distribution (cluster), k, is given as

$$z_{ik} = \frac{\xi_k N(x_i|\mu_k, \Sigma_k)}{\sum_{j=1}^{s} \xi_j N(x_i|\mu_j, \Sigma_j)}$$

And the expectation of the penalized log-likelihood is ---------- 3.29

$$Q_d(X|\mu, \Sigma, \xi) = \sum_{i=1}^{L}\sum_{k=1}^{d} z_{i,k}^{(0)} ln(\xi_k N(x_i|\mu_k, \Sigma_k)) - \Psi \quad \text{---------- 3.30}$$

Where Ψ is the penalization term.

Once we have the expected log-likelihood, we optimize to ensure data point is in the correct cluster in the M-step

- The M-step maximizes Q (expected log-likelihood) over μ_k, Σ_k and ξ_k

$$\frac{\partial Q}{\partial \mu_k} = 0 \ \ and \ \ \frac{\partial Q}{\partial \Sigma_k} = 0 \quad \text{---------- 3.31}$$

with

$$\mu_k^{(1)} = \frac{\sum_{i=1}^{L} z_{i,k}^{(0)} x_i}{\sum_{i=1}^{L} z_{i,k}^{(0)}} \quad \text{---------- 3.32}$$

$$\Sigma_k^{(1)} = \frac{\sum_{i=1}^{L} z_{i,k}^{(0)} \left(x_i - \mu_k^{(1)}\right)\left(x_i - \mu_k^{(1)}\right)^T}{\sum_{i=1}^{L} z_{i,k}^{(0)}} \quad \text{---------- 3.33}$$

$$\xi_k^{(0)} = \frac{1}{L}\sum_{i=1}^{L} z_{i,k}^{(0)} \quad \text{---------- 3.34}$$

Hence, we bin each data point into its optimum cluster, with its own distribution, however many clusters we choose to bin the randomly generated data set into.

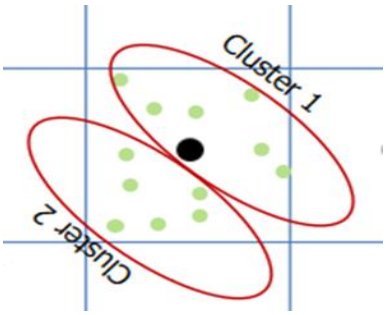


Figure 21: Grid cell depicting binning of permeability multipliers into clusters

- We further calculate and use the centroid of each cluster as an average for each cluster – depicted by the red data points in Figure 14. The geometric mean of these centroids becomes the optimum permeability multiplier for the contributing unobserved grid cell – depicted by the cyan data point in Figure 14.

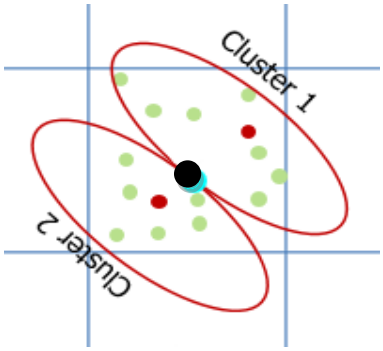


Figure 22: Grid cell depicting cluster centroid and geometric mean of centroids as grid cell permeability multiplier

- Given the approach of modifying the permeability multiplier on a cell-by-cell bases, we are confronted with some challenges as a result. If we have the permeability multiplier of a grid cell that is less than 1, consequently making it tighter and more difficult for fluids to pass through the grid cell, the fluid flow will be diverted to other nearby cell or cells. This means that the modification of each cells potentially has a domino effect on other nearby cells as well as on other cells that are further away through which the fluid that was originally supposed to pass through the modified cell are now passing. The fluids may very well be diverted to other producers other than the original producer they would have flowed into. This would have an unintended consequence of reducing the error in one producer which unintentionally increasing the error in another producer. Similar to the traveling sales man problem, only that in this situation, we have multiple traveling sales men (one for each producer), the challenge is to find permeability multipliers that will minimize the errors of all the producers at the same time, as depicted below.

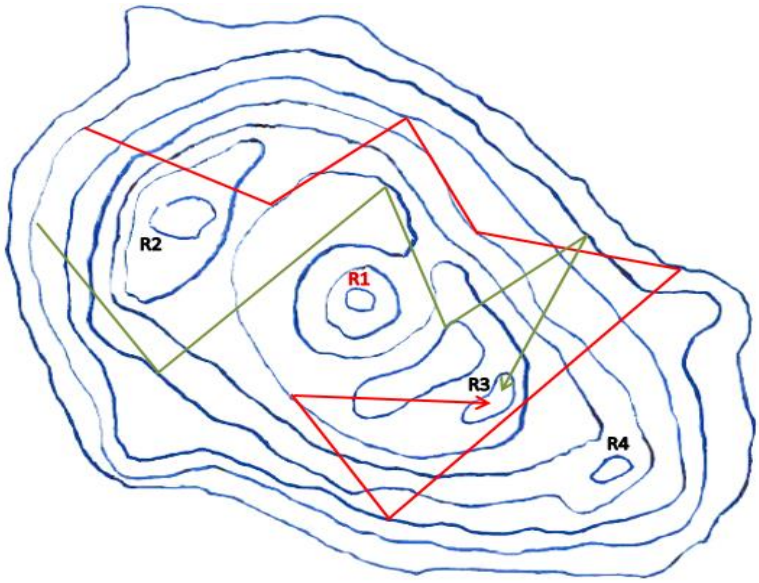


Figure 23: Non-Uniqueness of Solution and Non-Monotonic Convergence Nature of History Matching Method

- Also, the figure above shows four solutions, R1, R2, R3, and R4 with R1 being the true solution. With this method, we are sure to converge to a solution within some error threshold. However, as with all other history matching techniques and the inherent nature of history matching, there is no guarantee that we will converge to the true solution.

CHAPTER IV

RESULTS

Models

Both Normalized Error and EM Algorithm methods described in Chapter III were used to history match two standard models in the oil and gas industry. These two models are:

1. The Egg model
2. The SPE10 model

The outcome of the history marching methods was compared with the well know and established DOE method.

The Egg Model

The Egg model **[84]** is a two-phase heterogeneous dead Oil-Water system that consists of 8 injectors and 4 producers to make a total of 12 vertical wells in a 5-spot waterflood pattern. The picture below depicts the egg model

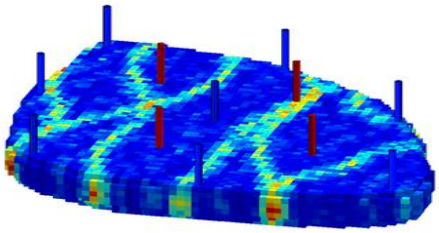


Figure 24: The Egg model – A geological ensemble for reservoir simulation (Reprinted from [85])

The system is of uniform porosity without capillary pressure considerations. The egg model consists of a total of 25,200 cells (active and inactive). There are 18,553 active cells with the inactive cells on the edges which makes the egg model depart from a regular layered cake model, taking the shape of an egg – hence its name. The egg model consists of high permeability streaks in a low permeability environment with permeabilities ranging from 97.6 md to 7000 md

The reservoir data and dimensions are as listed in Table 1 below.

Variable	Value
Nx	60
Ny	60
Nz	7
Number of Active Cells	18,553
Δz (grid block height)	4 m
Δx, Δy (Grid block length, width)	8 m
Porosity	0.2
Oil Compressibility	1e-10 Pa^{-1}
Rock Compressibility	0 Pa^{-1}
Water Compressibility	1e-10 Pa^{-1}
Oil Dynamic Viscosity	5e-3 Pa S
Water Dynamic Viscosity	1e-3 Pa S

Relative Permeability End-point for Oil	0.8
Relative Permeability End-point for Water	0.75
Corey Exponent, Oil	4
Corey Exponent, Water	3
Residual-oil saturation	0.1
Connate-water saturation	0.2
Capillary pressure	0 Pa
Initial reservoir pressure (top layer)	40e+6 Pa
Initial water saturation	0.1
Water injection rates, per well	79.5 m^3/day
Production well bottom-hole pressures	39.5e+6 Pa
Well-bore radius	0.1 m

Table 1: Egg model data and dimensions [85]

SPE10 Model

The SPE10 model **[86]** is a two-phase heterogeneous dead Oil-Water system that consists of 9 injectors and 4 producers to make a total of 13 vertical wells in a 5-spot waterflood pattern with the producers at the center of the system and the injectors on the outer of the system surrounding the producers.

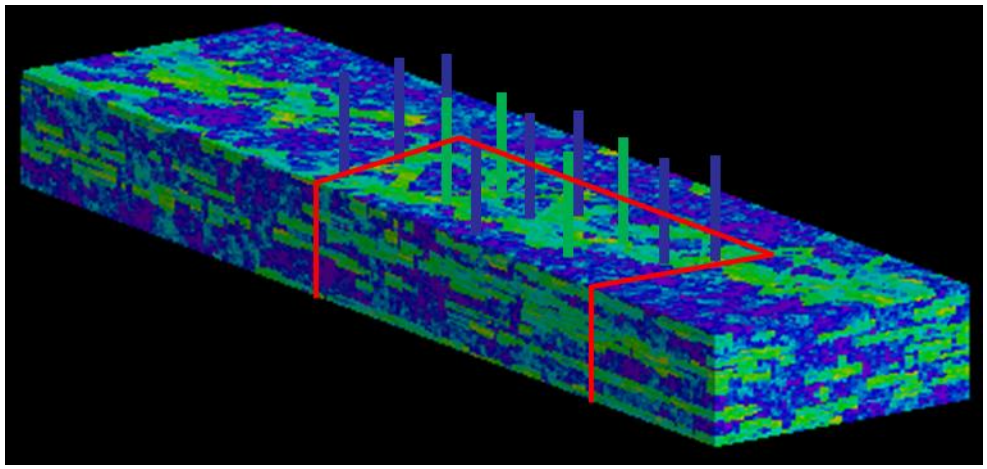


Figure 25: SPE10 Comparative Study showing the subsection and well locations for research study. (Reprinted from [86])

We have used parameters values similar to the Egg model for ease of comparison. The system is of uniform porosity without capillary pressure considerations. The SPE10 model consists of a total of 66,000 cells, all of which are active. The section of the SPE10 model we have utilized for this research is highly heterogeneous with its permeabilities ranging from 10 md to 2000 md

The reservoir data and dimensions are as listed in Table 2 below.

Variable	Value
Nx	60
Ny	110
Nz	10
Number of Active Cells	66,000

Δz (grid block height)	4 m
Δx, Δy (Grid block length, width)	8 m
Porosity	0.2
Oil Compressibility	1e-10 Pa^{-1}
Rock Compressibility	0 Pa^{-1}
Water Compressibility	1e-10 Pa^{-1}
Oil Dynamic Viscosity	5e-3 Pa S
Water Dynamic Viscosity	1e-3 Pa S
Relative Permeability End-point for Oil	0.8
Relative Permeability End-point for Water	0.75
Corey Exponent, Oil	4
Corey Exponent, Water	3
Residual-oil saturation	0.1
Connate-water saturation	0.2
Capillary pressure	0 Pa
Initial reservoir pressure (top layer)	40e+6 Pa
Initial water saturation	0.1
Water injection rates, per well	79.5 m^3/day
Production well bottom-hole pressures	39.5e+6 Pa
Well-bore radius	0.1 m

Table 2: SPE10 model data and dimensions used

History Matching Results

Given that all other variables remain constant as indicated in the respective tables above, we undertook the task of calculating the permeability distribution with the new method so as to match the oil and water production rates. As with any history match, it is important to note that the match oil and water production rates are not unique to one permeability distribution. The permeability distribution we arrive at is just one realization out of so many that will match the historical data. The history match for both models go out for 120 months with oil and water production rates reported monthly.

The Egg Model History Matching Result

For the Egg model **[84]**, our error threshold limit was set at 2% i.e. a match was not considered successful until the calculated error of all of the producers became less than or equal to 2% in the same loop run. In other words, if 3 of the 4 producers had calculated errors less than 2% but the 4^{th} producer turns out to have a calculated error greater than 2%, this would not be considered a match. All calculated errors for all the producers would have to be less than the 2% to be acceptable as a match.

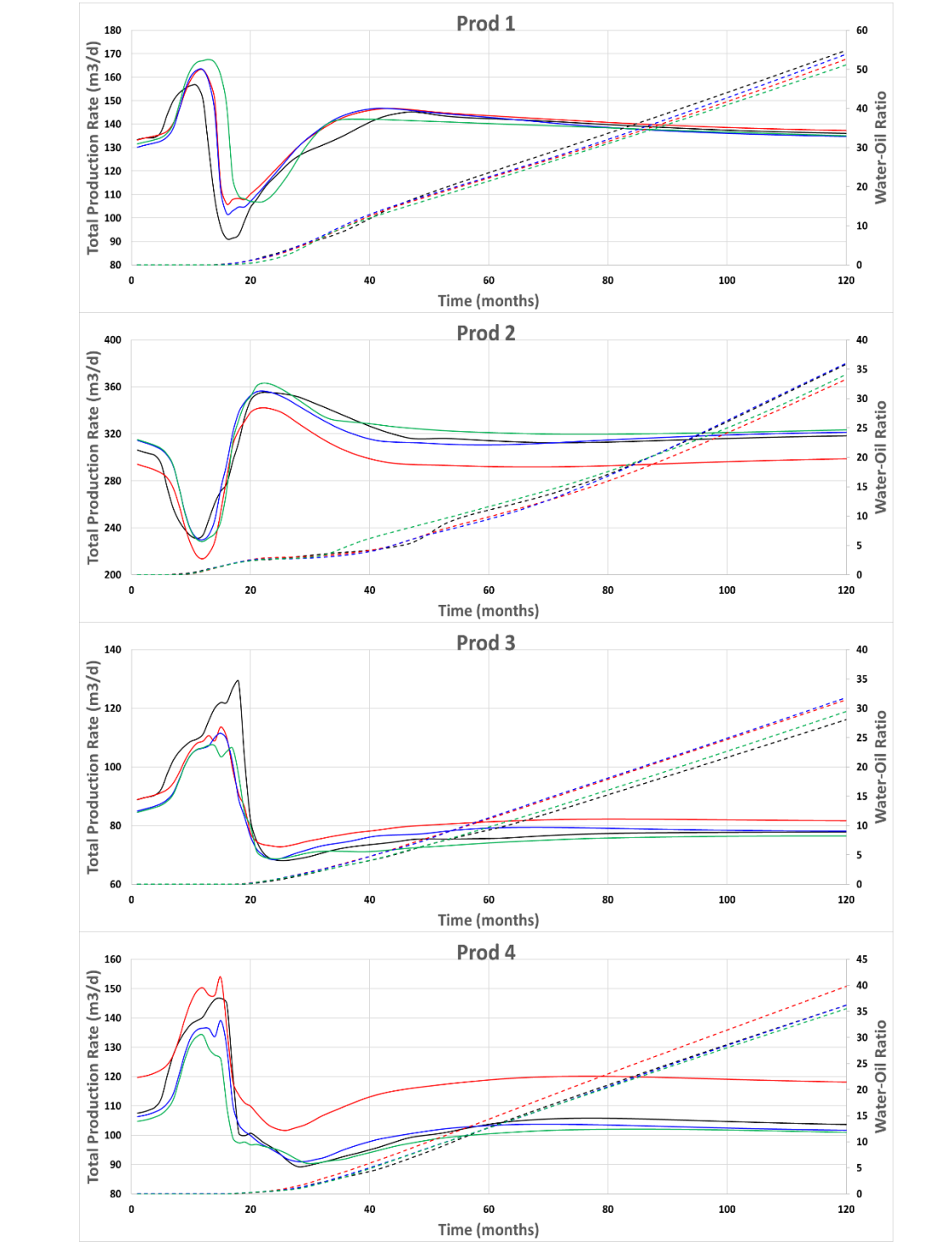


Figure 26: Comparison between the history matching methods the normalization method and the EM algorithm for the Egg Model, both starting with an initial guess using log-normal Kriging

Figure 26 above shows the history matches achieved for each of the four producers in the Egg model. In Figure 26, the black lines, solid and dash, are the historical total production and Water-Oil ratio, respectively. The red lines, solid and dash, are the total production and Water-Oil ratio respectively, calculated by running the initial permeability guess derived using log-normal Kriging in the Egg model. The green lines, solid and dash, are the total production and Water-Oil ratio respectively, calculated by running the permeability derived using the Normalized Error method in the Egg model. And, the blue lines, solid and dash, are the total production and Water-Oil ratio respectively, calculated by running the permeability derived using the EM Algorithm method in the Egg model.

In Figure 26, we see that we are able to get history matches closer to the historical data than where we initially started (initial log-normal Kriging guess), due to the implementation of the Normalized Error and EM Algorithm methods.

The result of the history match for both methods (permeability multipliers derived from normalized error without the EM algorithm and the application of the EM algorithm to the permeability multipliers derived from the normalized error) were compared to the history match using the DOE method and are shown in Figure 27.

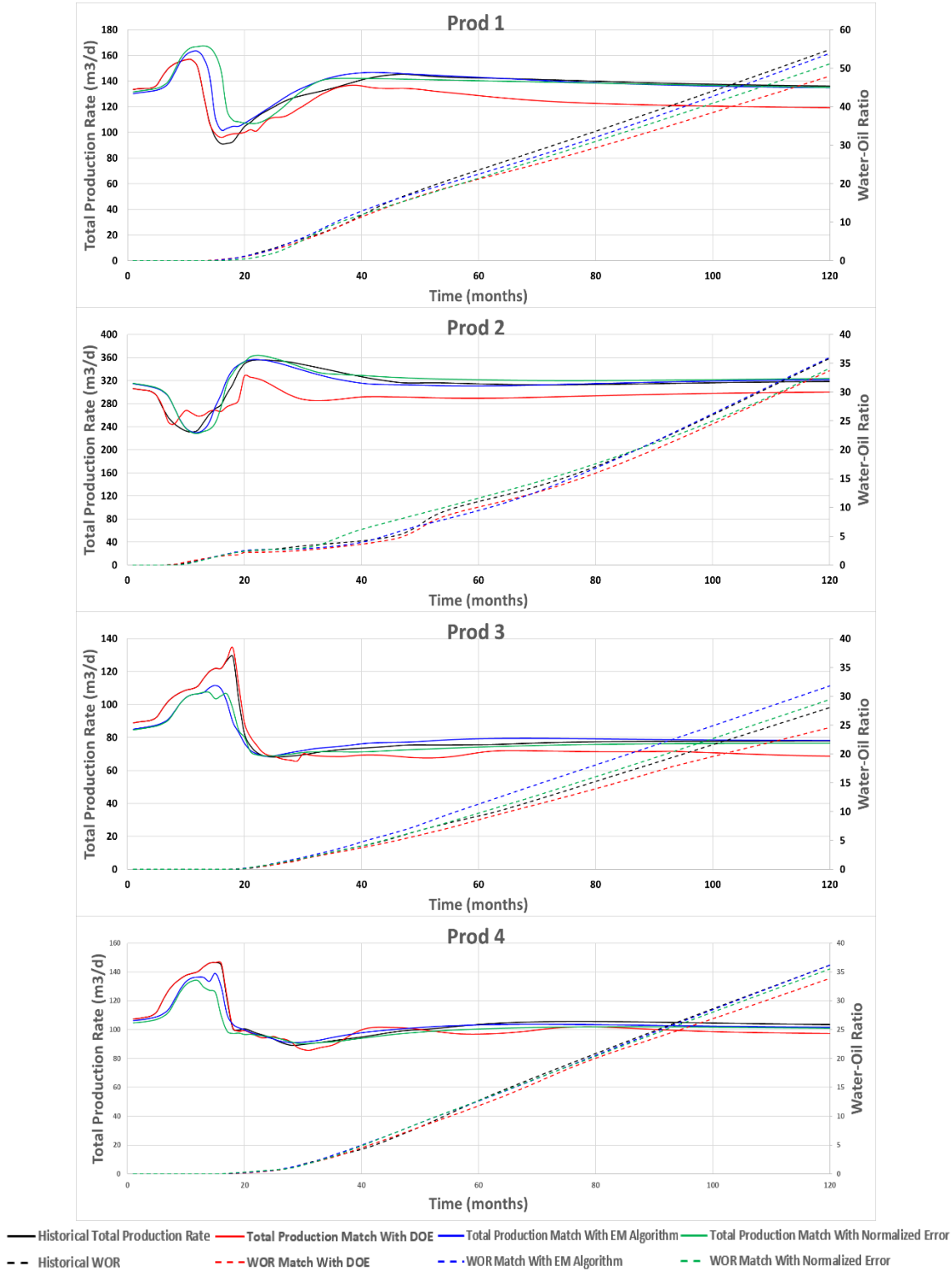


Figure 27: Comparison between the history matching methods of Experimental Design, the normalization method and the EM algorithm for the Egg model

In Figure 27, the black, green and blue lines are as described previously for Figure 26. However, the red line, solid and dash, are now the total production and Water-Oil ratio respectively, calculated by running the permeability derived using the DOE method in the Egg model.

In Figure 27, we see that we are able to get comparable history matches of the historical data using DOE when compared with history match done by the Normalized Error and EM Algorithm methods.

For the Egg model, the permeability distribution at each of the history matching process is shown for each layer. The permeability distribution for each layer of the EM Algorithm history match looks somewhat different from the true permeability distribution. This further confirms the non-uniqueness of history matching in general as well as the non-uniqueness of the new the EM Algorithm method at the center of this research.

In Figures 4.5 through 4.11, we show the permeability distribution of each layer of the Egg model as we go through the implementation of both the Normalized Error method and the EM Algorithm method. Each figure shows the permeability distribution of each layer and has four quadrants. In each figure (each layer), the top left quadrant of each figure shows the permeability distribution of sparse data with 95% missing permeability data; the top right quadrant of each figure shows the permeability distribution of the initial guess derived by applying log-normal Kriging to the sparse data shown in the top left quadrant; the bottom left quadrant of each figure shows the resulting permeability distribution derived by applying the EM Algorithm to the

permeabilities calculated with log-normal Kriging shown in the top right quadrant; the bottom right quadrant of each figure shows the true permeability distribution for that layer.

We get better results by we apply the EM Algorithm method to the log-normal Kriging output due to the cell-by-cell EM Algorithm optimization of the permeability multiplier used to update the field-wide permeabilities.

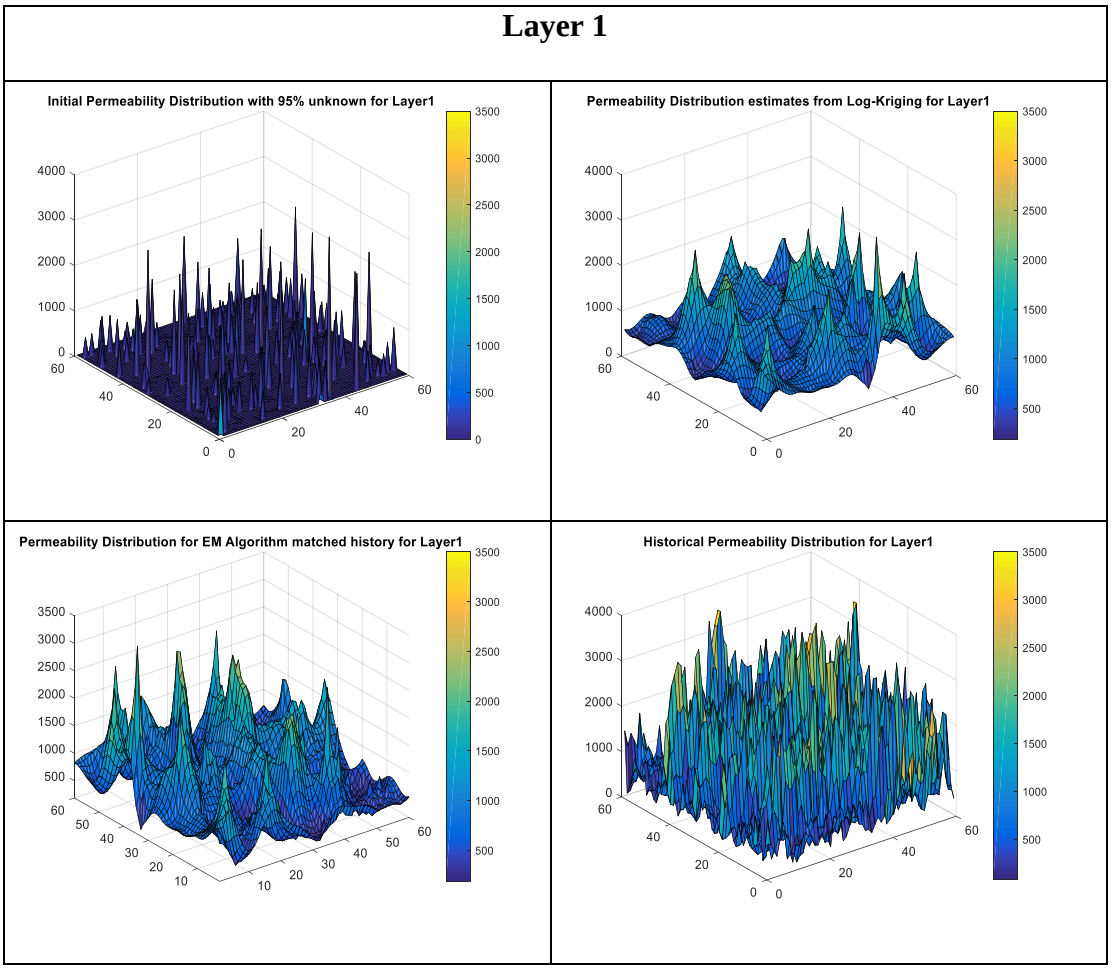


Figure 28: Permeability distribution at different workflow steps in comparison to the true permeability distribution Layer 1 of the Egg model

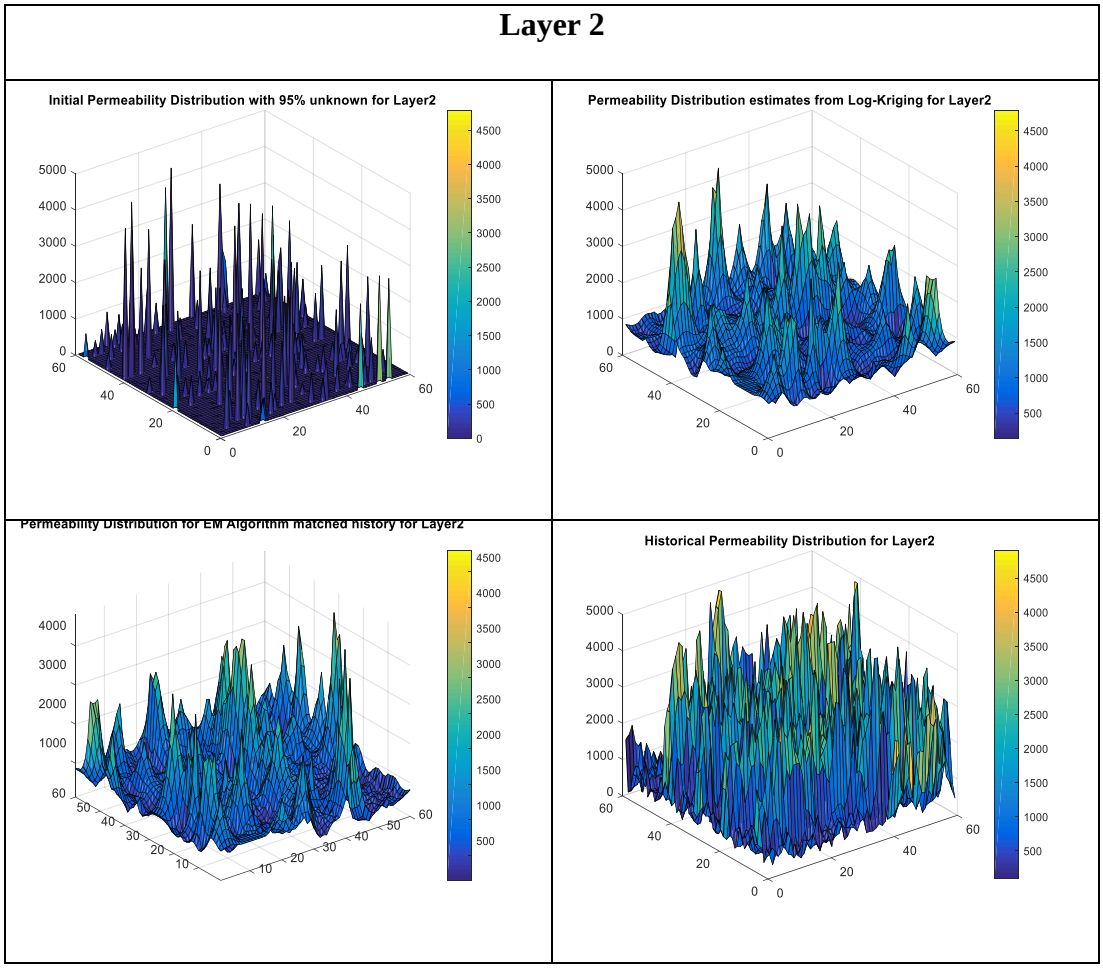


Figure 29: Permeability distribution at different workflow steps in comparison to the true permeability distribution Layer 2 of the Egg model

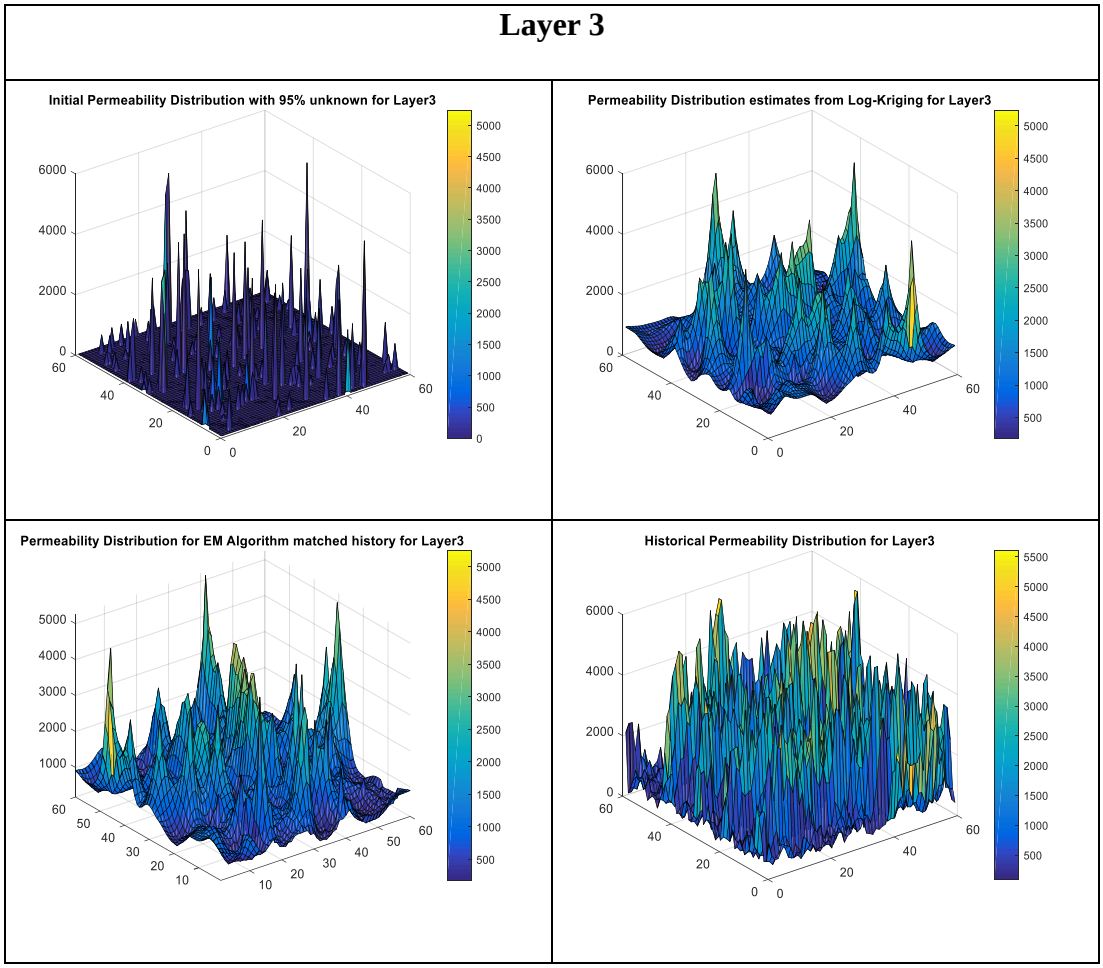


Figure 30: Permeability distribution at different workflow steps in comparison to the true permeability distribution Layer 3 of the Egg model

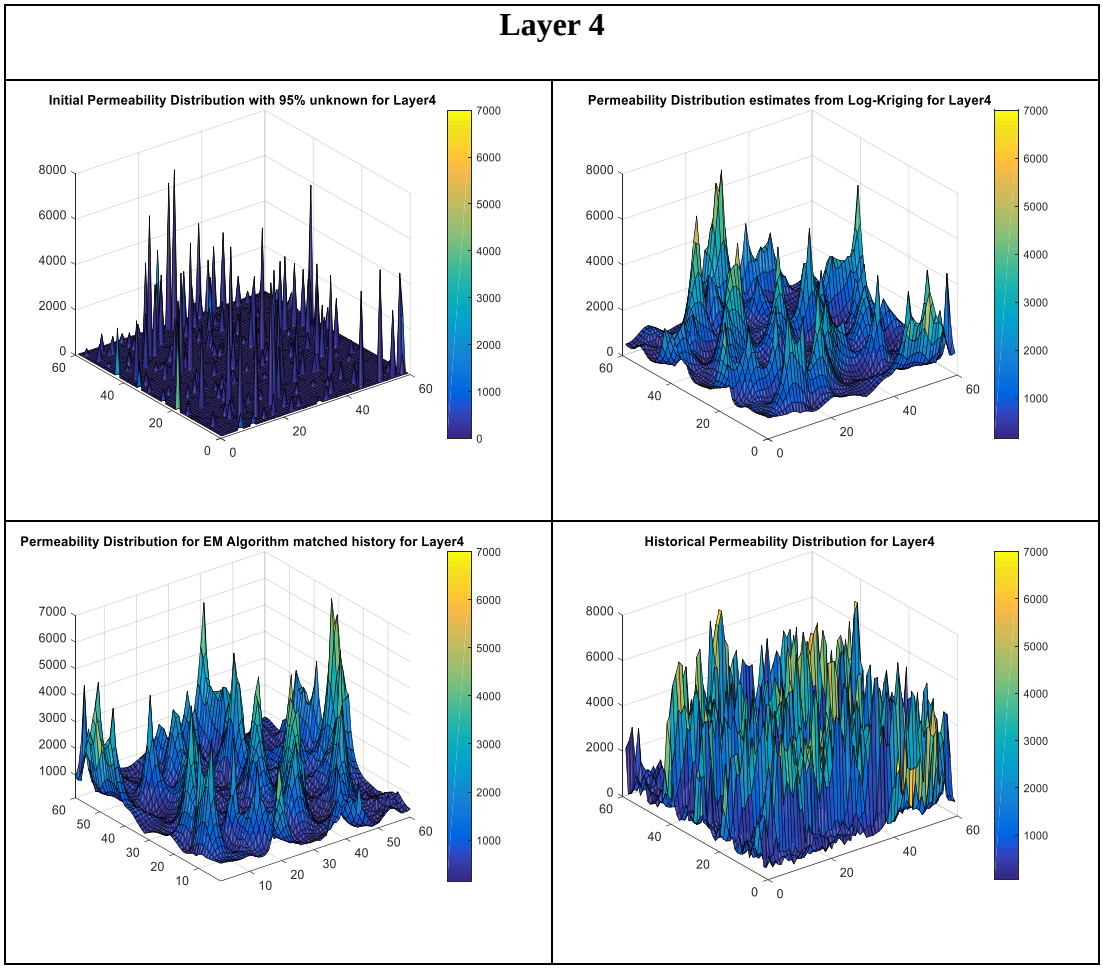


Figure 31: Permeability distribution at different workflow steps in comparison to the true permeability distribution Layer 4 of the Egg model

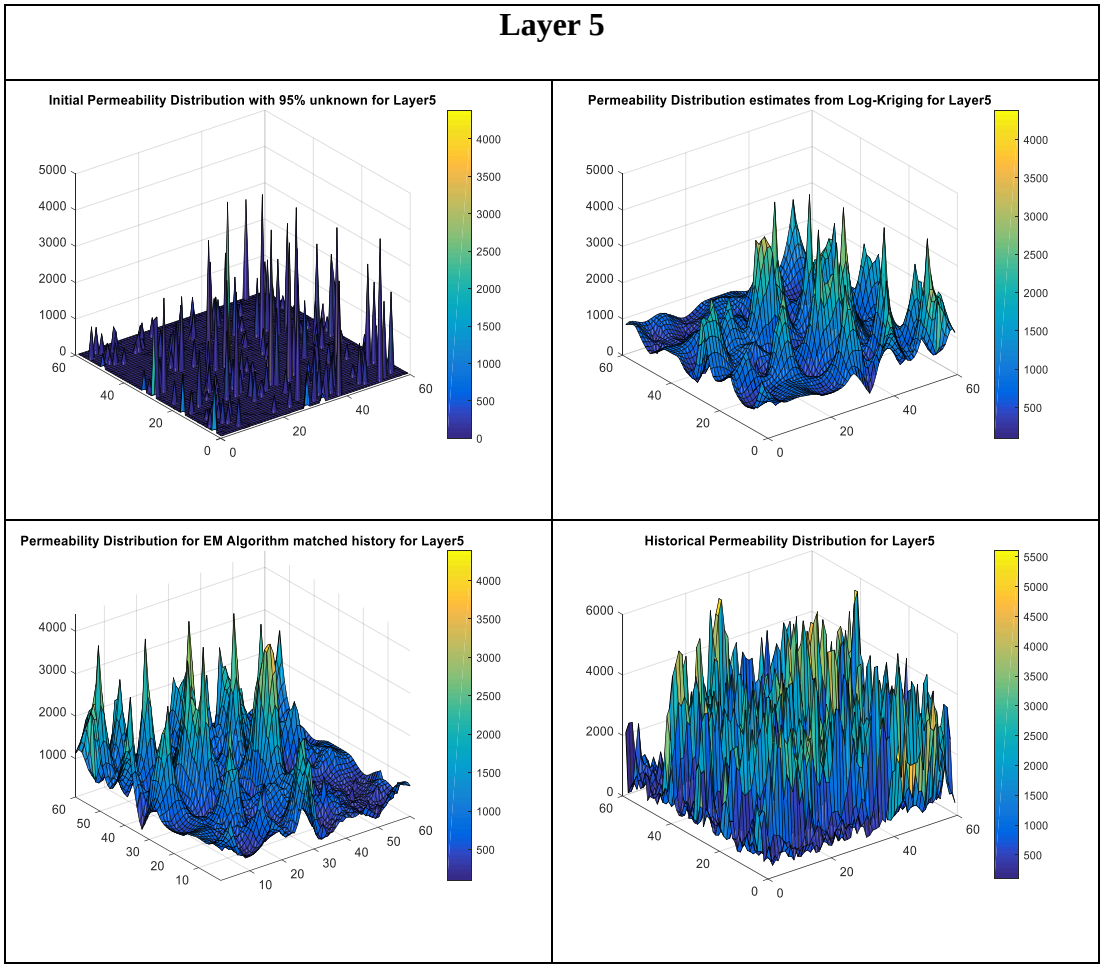


Figure 32: Permeability distribution at different workflow steps in comparison to the true permeability distribution Layer 5 of the Egg model

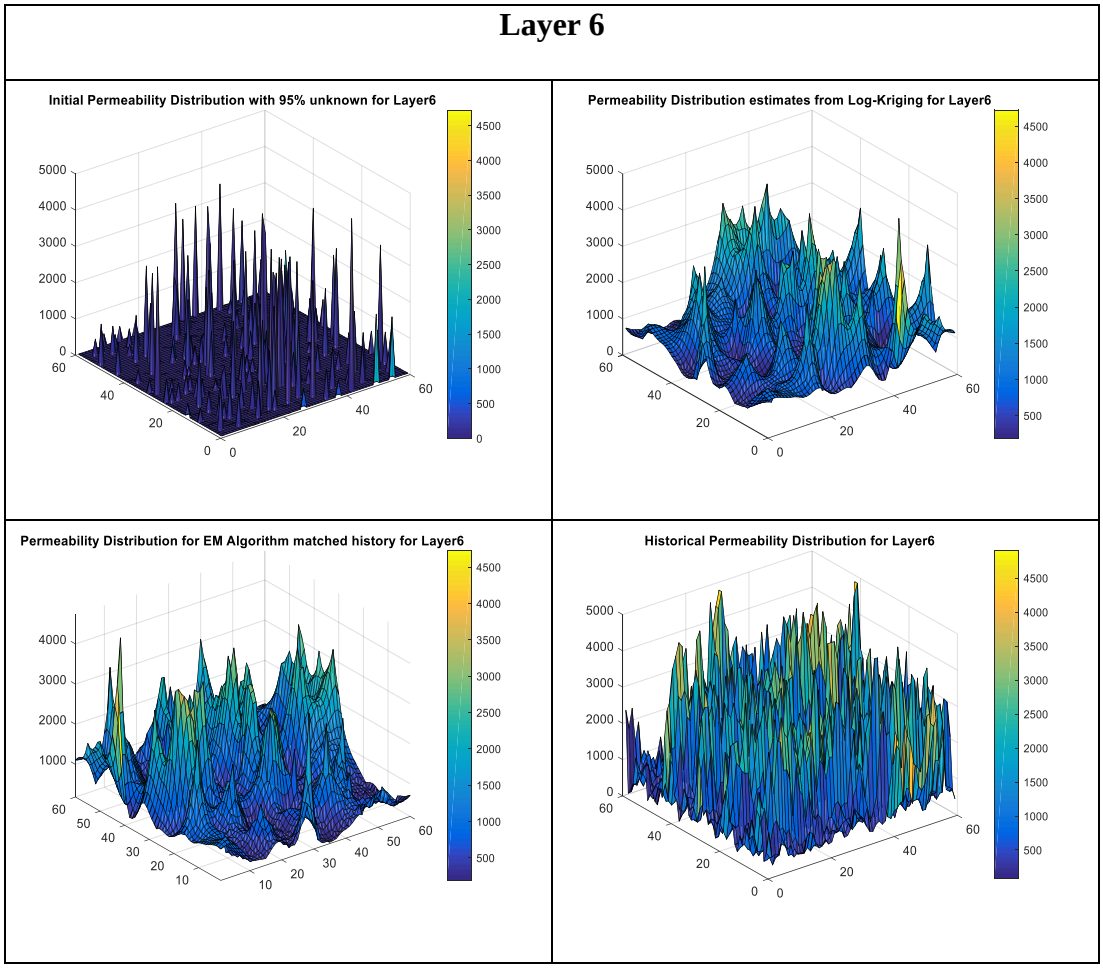


Figure 33: Permeability distribution at different workflow steps in comparison to the true permeability distribution Layer 6 of the Egg model

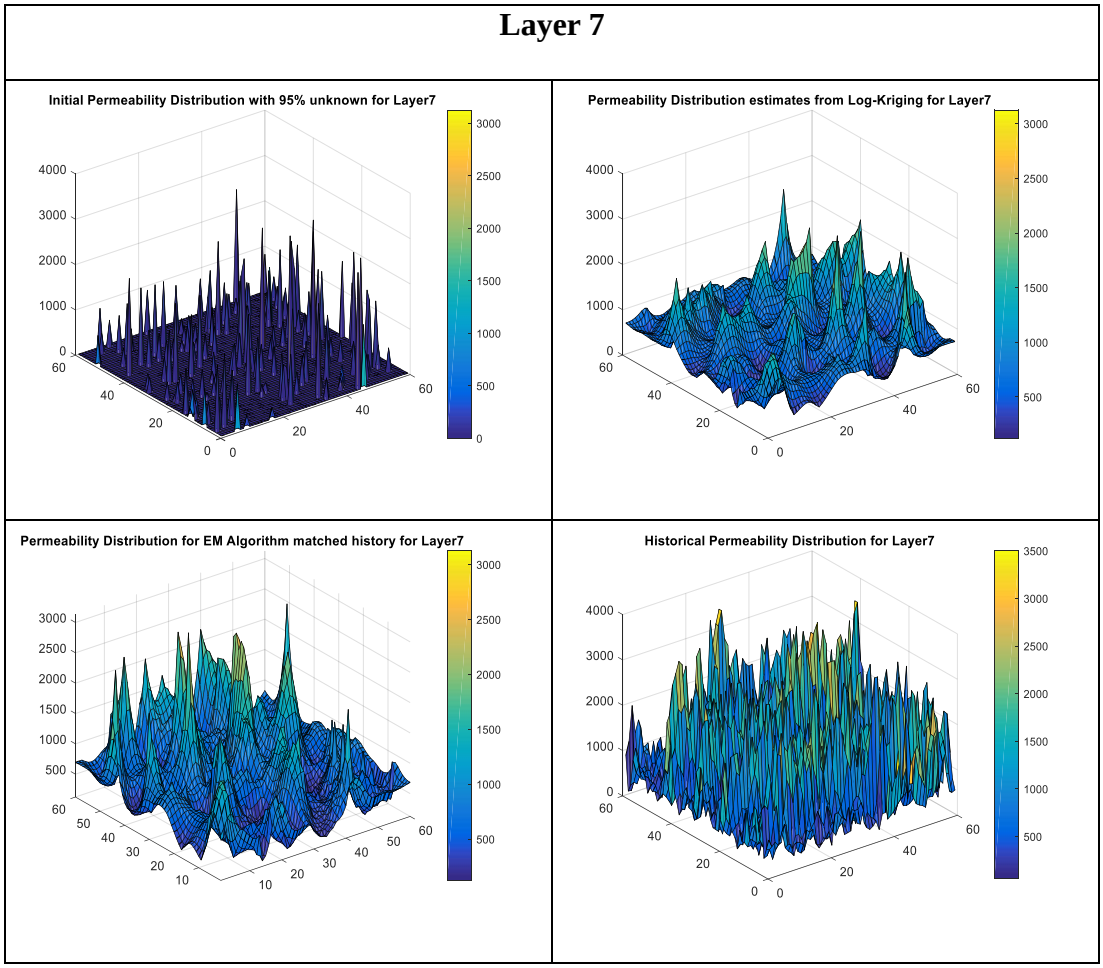


Figure 34: Permeability distribution at different workflow steps in comparison to the true permeability distribution Layer 7 of the Egg model

SPE10 Model History Matching Result

For the SPE10 model **[86]**, our error threshold limit was also set at 2%. Similar to the Egg model, a match was not considered successful until the calculated error of all of the producers became less than 2% in the same loop run. All calculated errors for all the producers would have to be less than the 2% to be acceptable as a match.

Below is the table showing the history matches achieved for each of the four producers in the SPE10 model. The first column contains the outcome of the run that was made with the initial permeability guess using log-normal Kriging. The second column contains the matches achieved by just using the permeability multipliers derived from normalized error without further application of the EM algorithm. The third column contains the matches achieved by applying the EM algorithm to the permeability multipliers derived from normalized error.

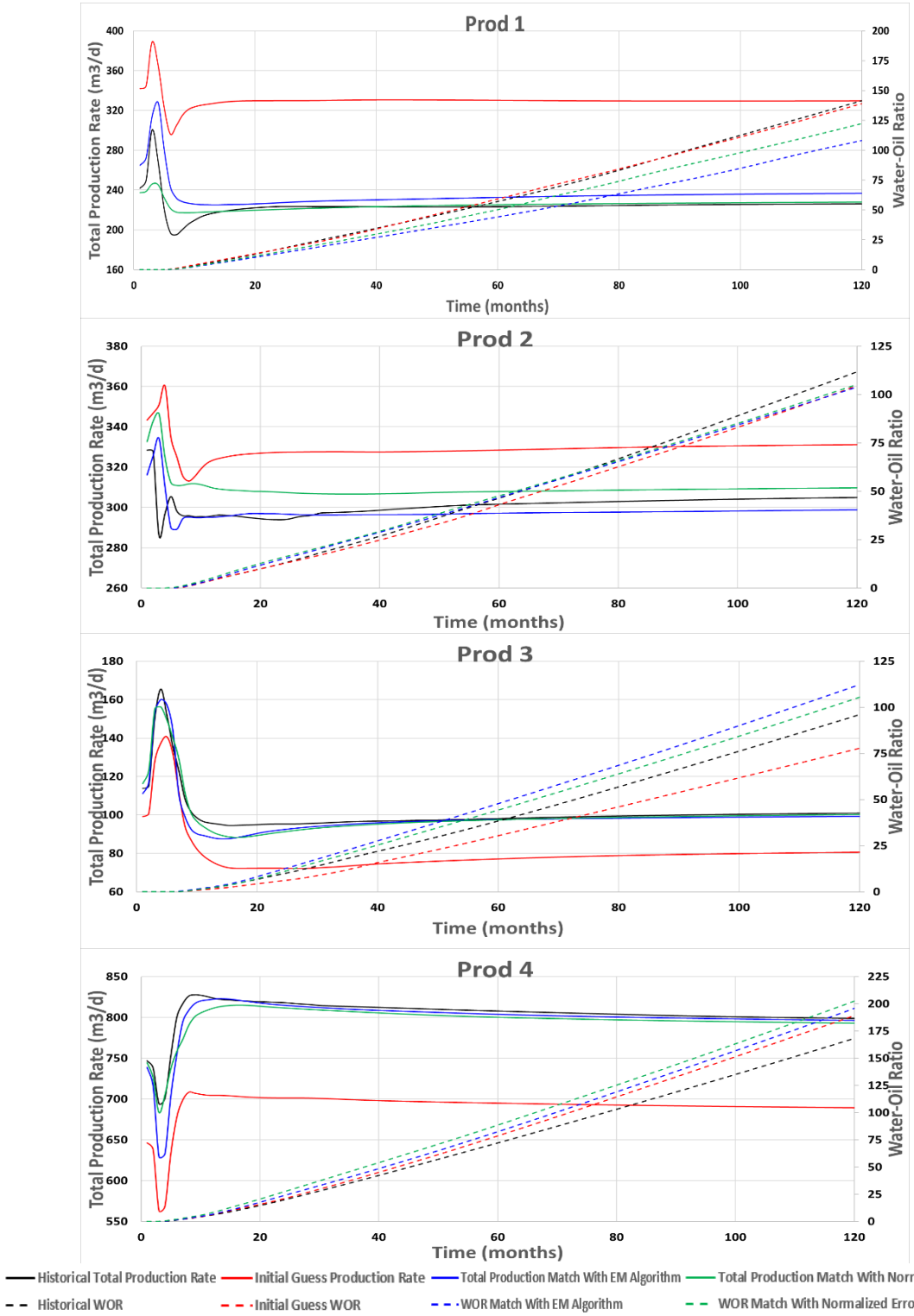


Figure 35: Comparison between the history matching methods the normalization method and the EM algorithm for the SPE10 Model, both starting with an initial guess using log-normal Kriging

Figure 35 above shows the history matches achieved for each of the four producers in the SPE10 model. Similar to Figure 26, in Figure 35, the black lines, solid and dash, are the historical total production and Water-Oil ratio, respectively. The red lines, solid and dash, are the total production and Water-Oil ratio respectively, calculated by running the initial permeability guess derived using log-normal Kriging in the Egg model. The green lines, solid and dash, are the total production and Water-Oil ratio respectively, calculated by running the permeability derived using the Normalized Error method in the Egg model. And, the blue lines, solid and dash, are the total production and Water-Oil ratio respectively, calculated by running the permeability derived using the EM Algorithm method in the Egg model.

In Figure 35, we see that we are able to get history matches closer to the historical data than where we initially started (initial log-normal Kriging guess), due to the implementation of the Normalized Error and EM Algorithm methods.

The result of the history match for both methods (permeability multipliers derived from normalized error without the EM algorithm and the application of the EM algorithm to the permeability multipliers derived from the normalized error) were result compared to the history match using the DOE method and are shown in Figure 36.

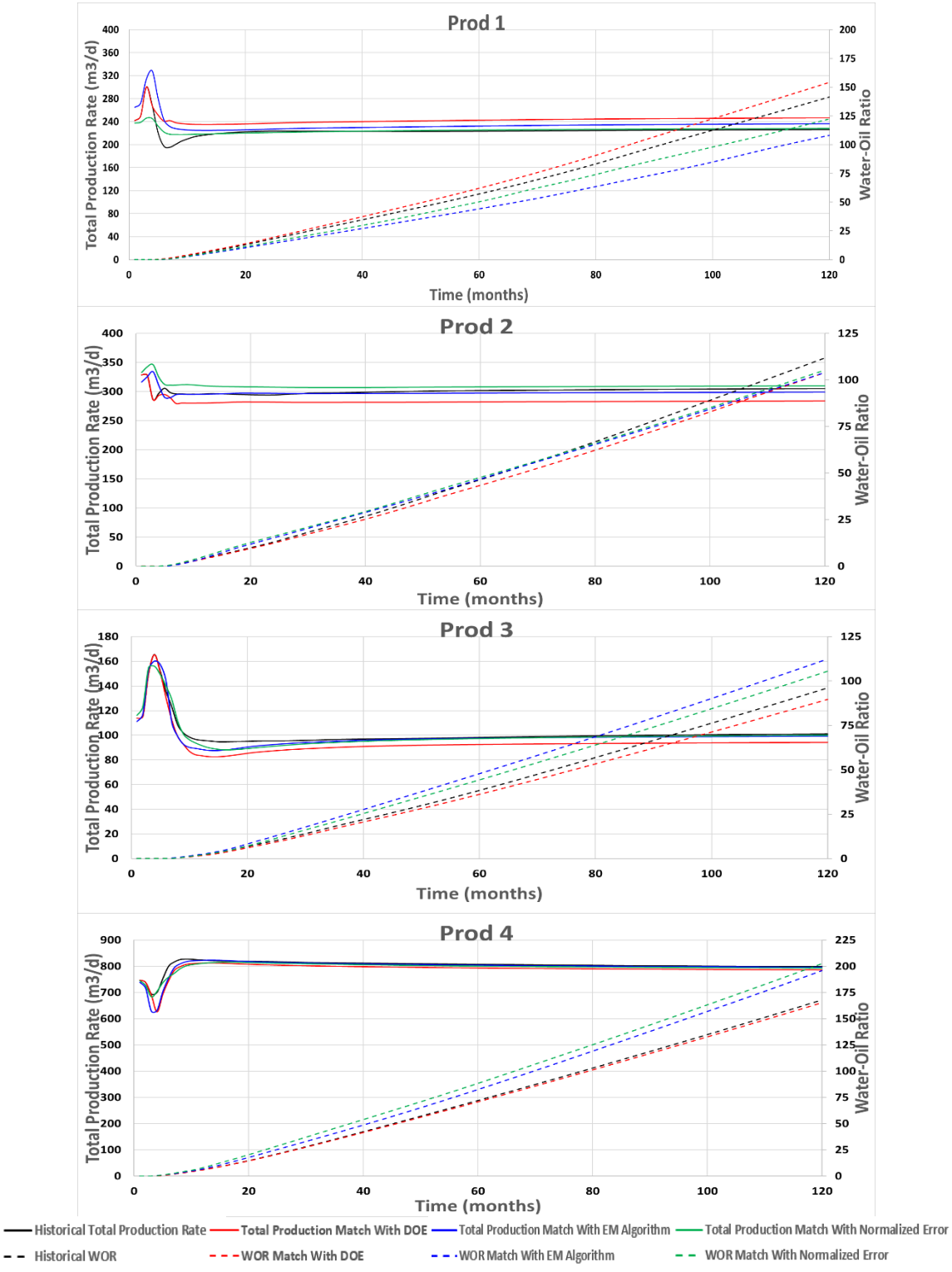


Figure 36: Comparison between the history matching methods of Experimental Design, the normalization method and the EM algorithm for the SPE10 model

In Figure 36, the black, green and blue lines are as described previously for Figure 35. However, the red line, solid and dash, are now the total production and Water-Oil ratio respectively, calculated by running the permeability derived using the DOE method in the SPE10 model.

In Figure 36, we see that we are able to get comparable history matches of the historical data using DOE when compared with history match done by the Normalized Error and EM Algorithm methods.

For the SPE10 model, the permeability distribution at each of the history matching process is shown for each layer. The permeability distribution for each layer of the EM Algorithm history match looks somewhat different from the true permeability distribution. This also confirms the non-uniqueness of history matching in general as well as the non-uniqueness of the new the EM Algorithm method at the center of this research.

Similar to Figures 28 through 34 above, in Figures 37 through 46, we show the permeability distribution of each layer of the Egg model as we go through the implementation of both the Normalized Error method and the EM Algorithm method. Each figure shows the permeability distribution of each layer and has four quadrants. In each figure (each layer), the top left quadrant of each figure shows the permeability distribution of sparse data with 95% missing permeability data; the top right quadrant of each figure shows the permeability distribution of the initial guess derived by applying log-normal Kriging to the sparse data shown in the top left quadrant; the bottom left quadrant of each figure shows the resulting permeability distribution derived by applying

the EM Algorithm to the permeabilities calculated with log-normal Kriging shown in the top right quadrant; the bottom right quadrant of each figure shows the true permeability distribution for that layer.

We get better results by we apply the EM Algorithm method to the log-normal Kriging output due to the cell-by-cell EM Algorithm optimization of the permeability multiplier used to update the field-wide permeabilities.

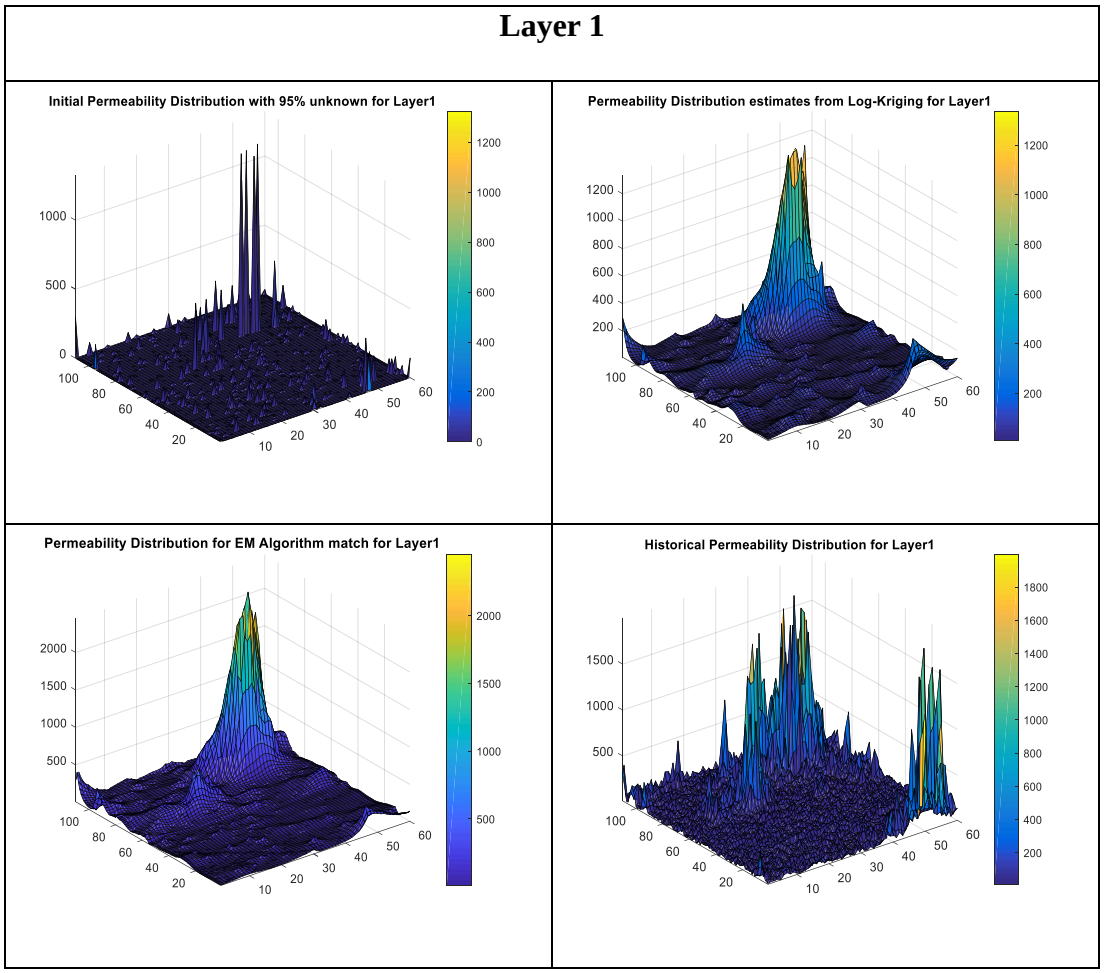


Figure 37: Permeability distribution at different workflow steps in comparison to the true permeability distribution Layer 1 of the SPE10 model

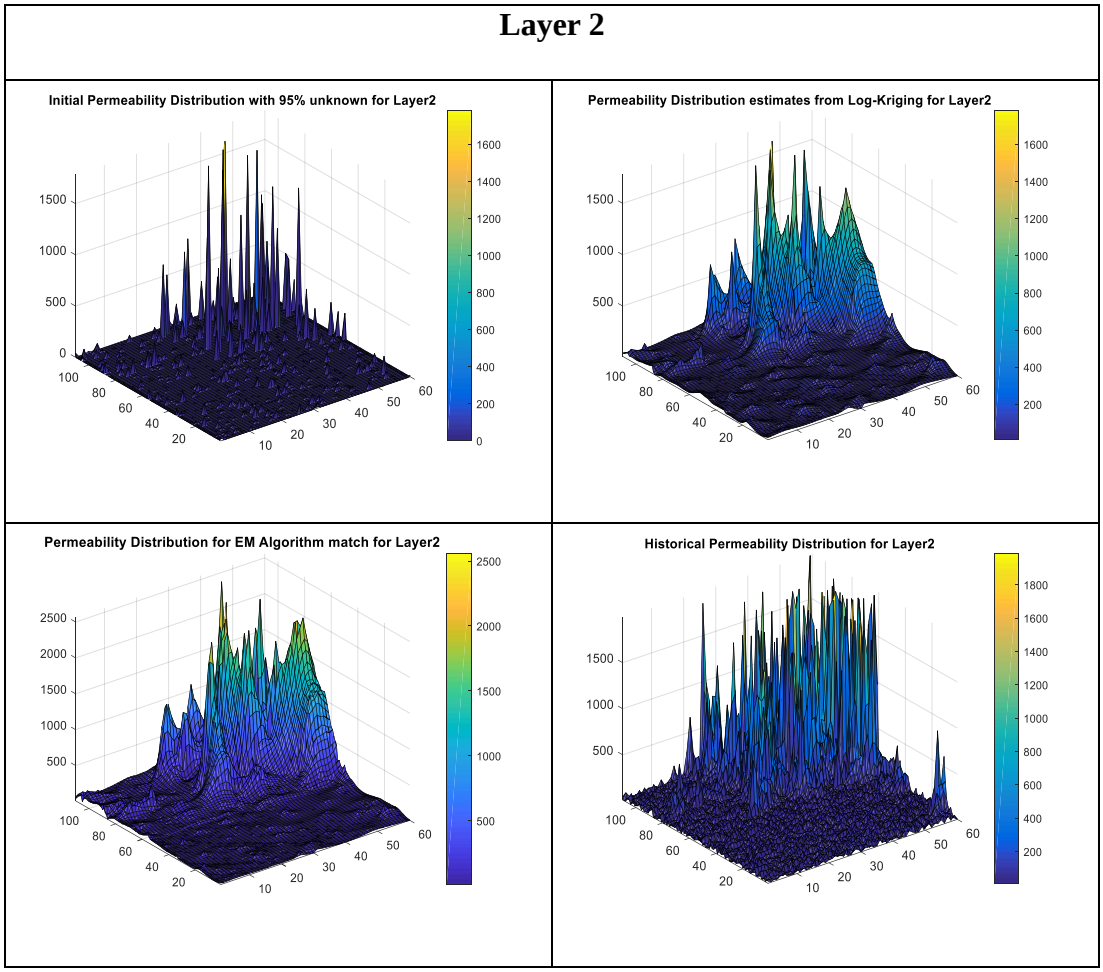


Figure 38: Permeability distribution at different workflow steps in comparison to the true permeability distribution Layer 2 of the SPE10 model

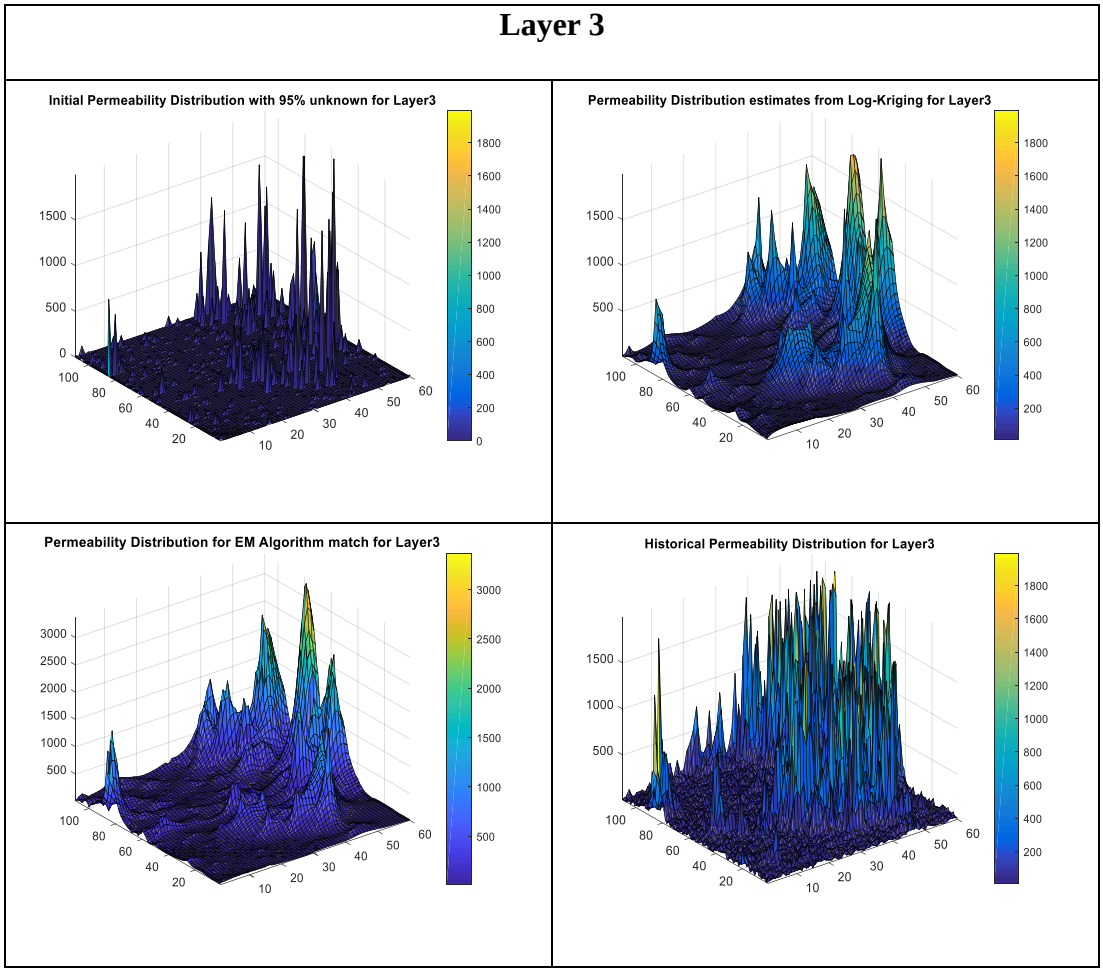


Figure 39: Permeability distribution at different workflow steps in comparison to the true permeability distribution Layer 3 of the SPE10 model

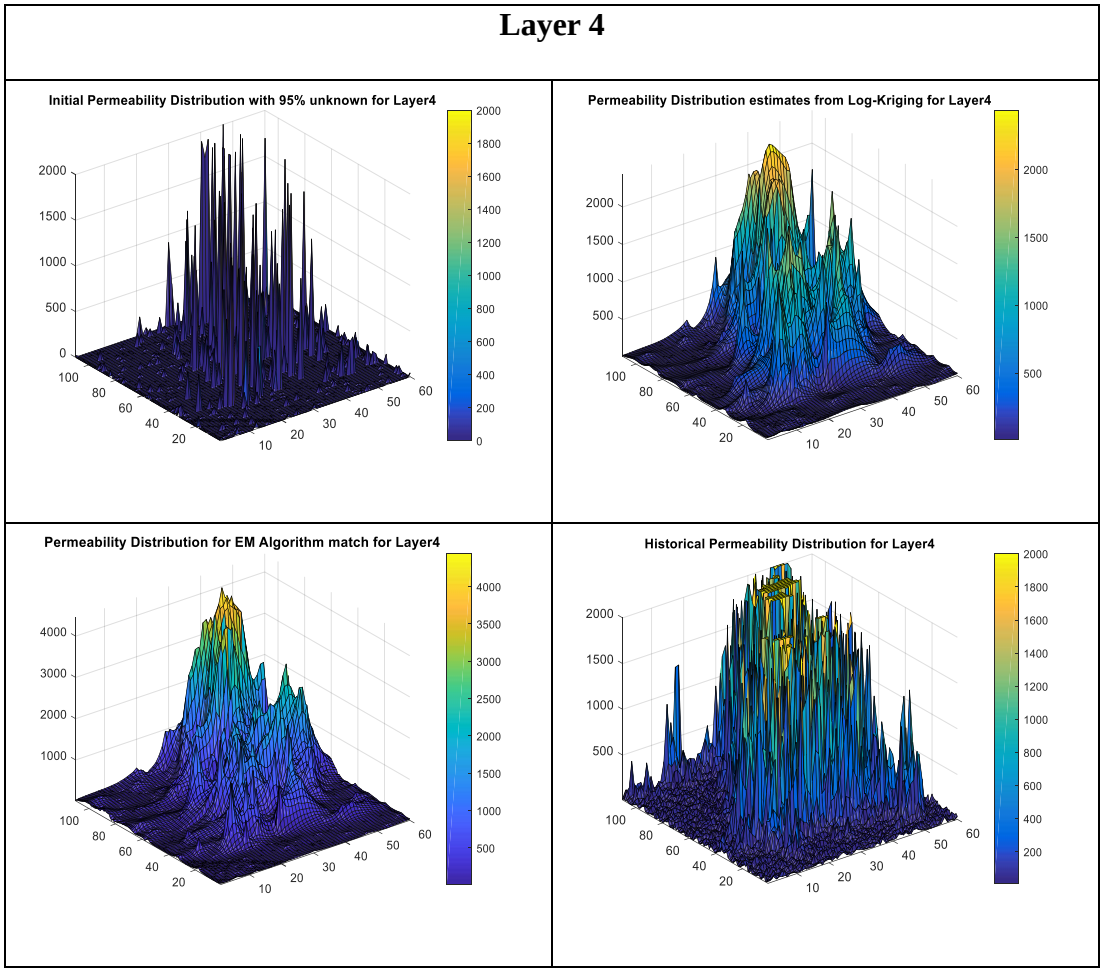


Figure 40: Permeability distribution at different workflow steps in comparison to the true permeability distribution Layer 4 of the SPE10 model

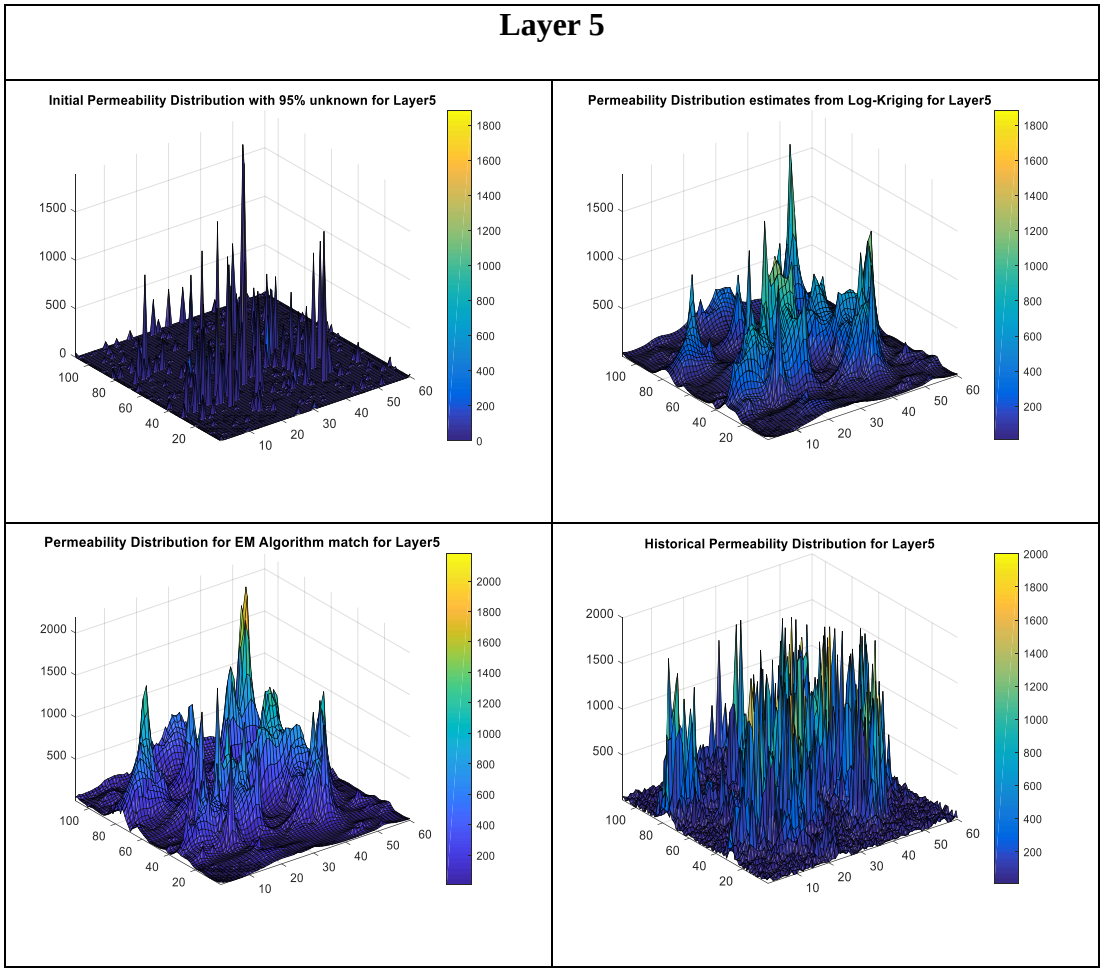


Figure 41: Permeability distribution at different workflow steps in comparison to the true permeability distribution Layer 5 of the SPE10 model

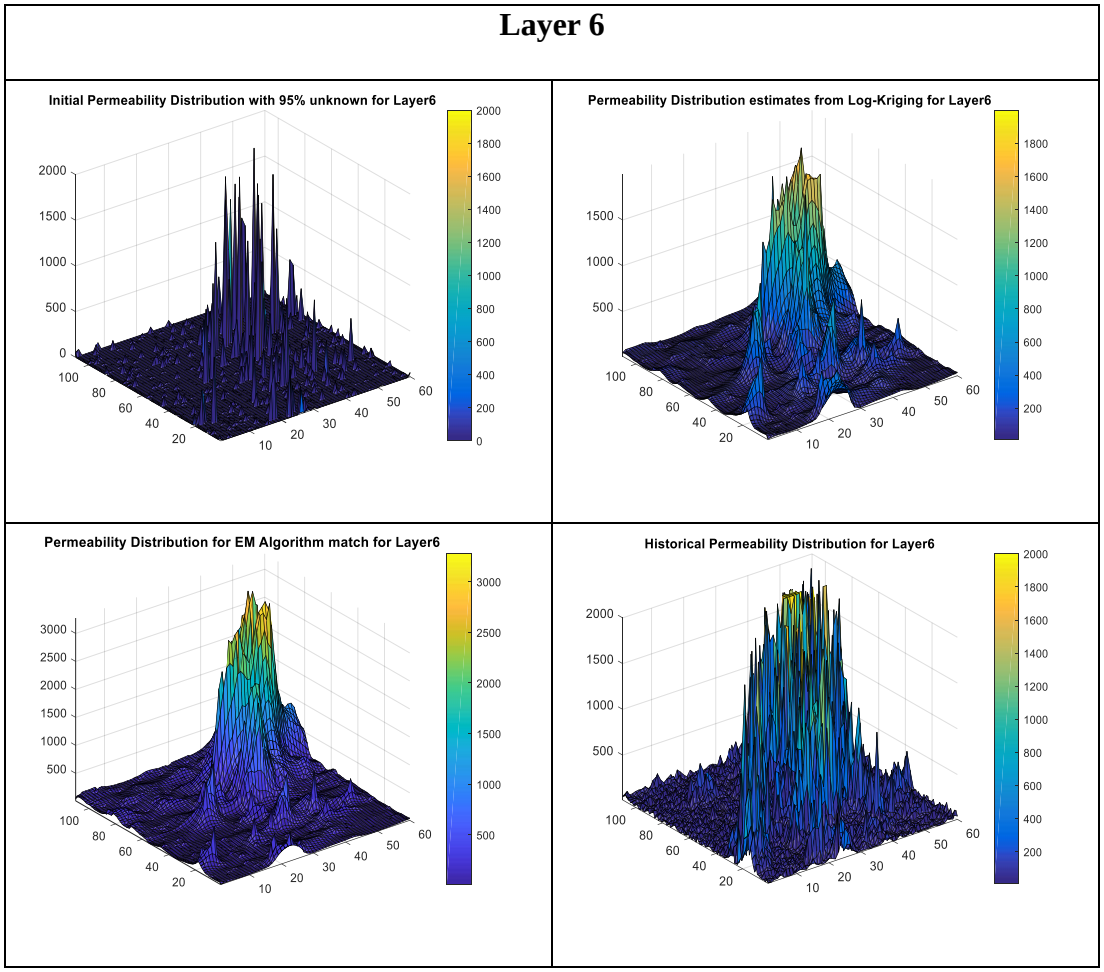


Figure 42: Permeability distribution at different workflow steps in comparison to the true permeability distribution Layer 6 of the SPE10 model

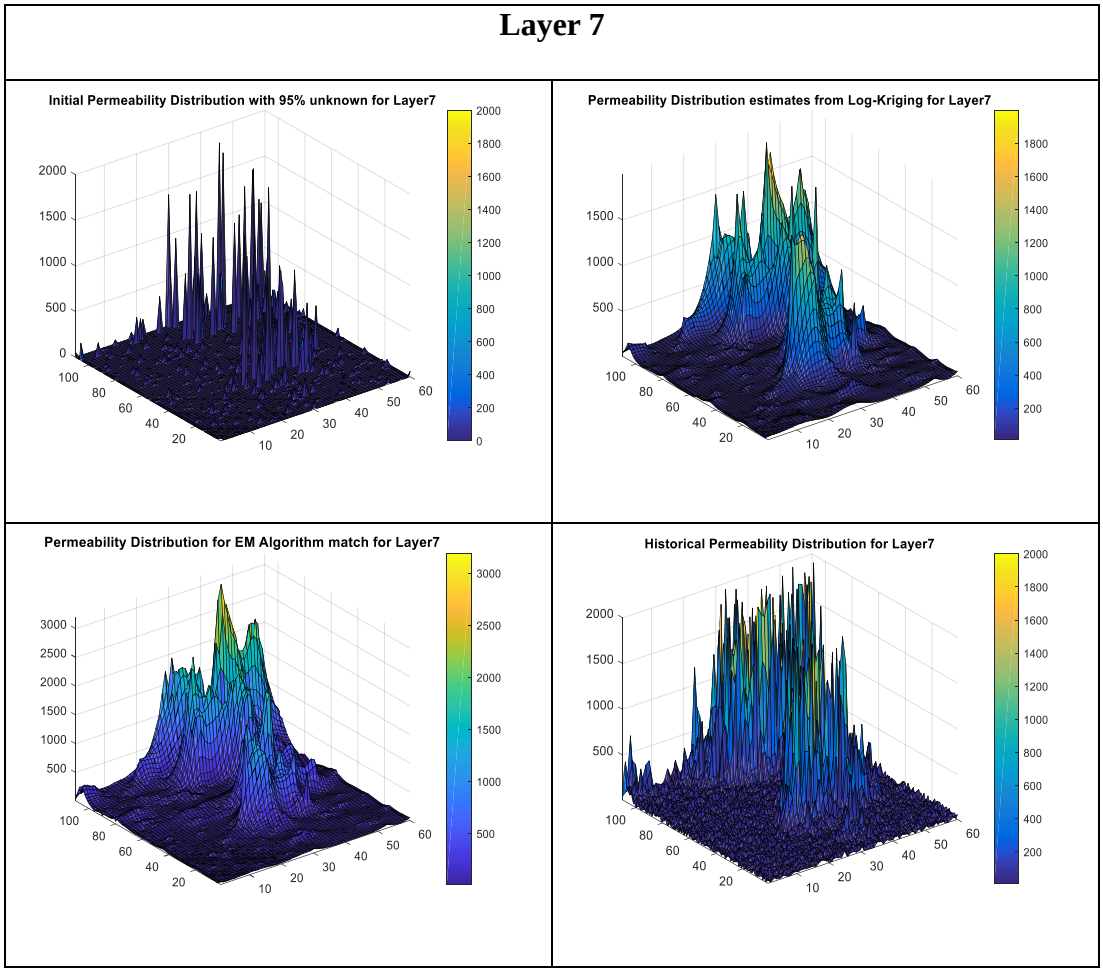


Figure 43: Permeability distribution at different workflow steps in comparison to the true permeability distribution Layer 7 of the SPE10 model

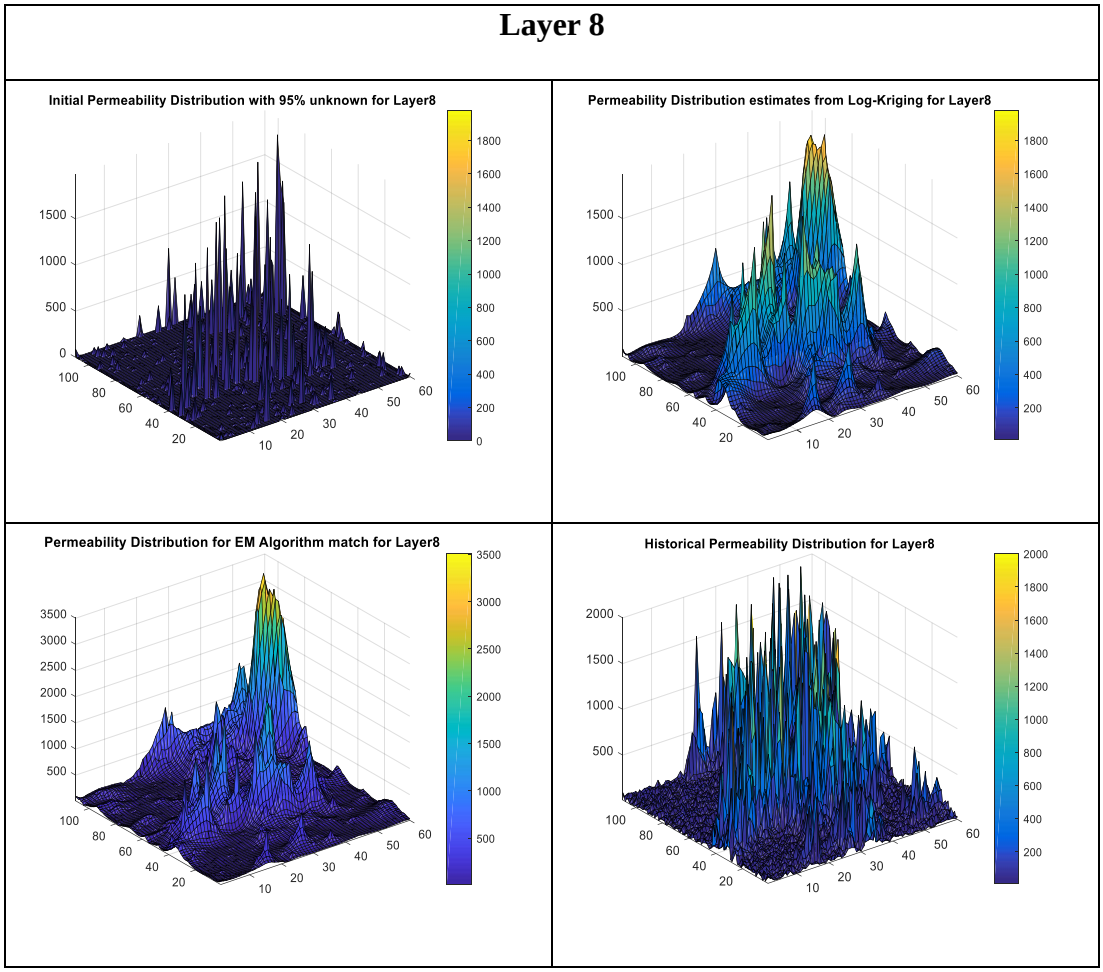


Figure 44: Permeability distribution at different workflow steps in comparison to the true permeability distribution Layer 8 of the SPE10 model

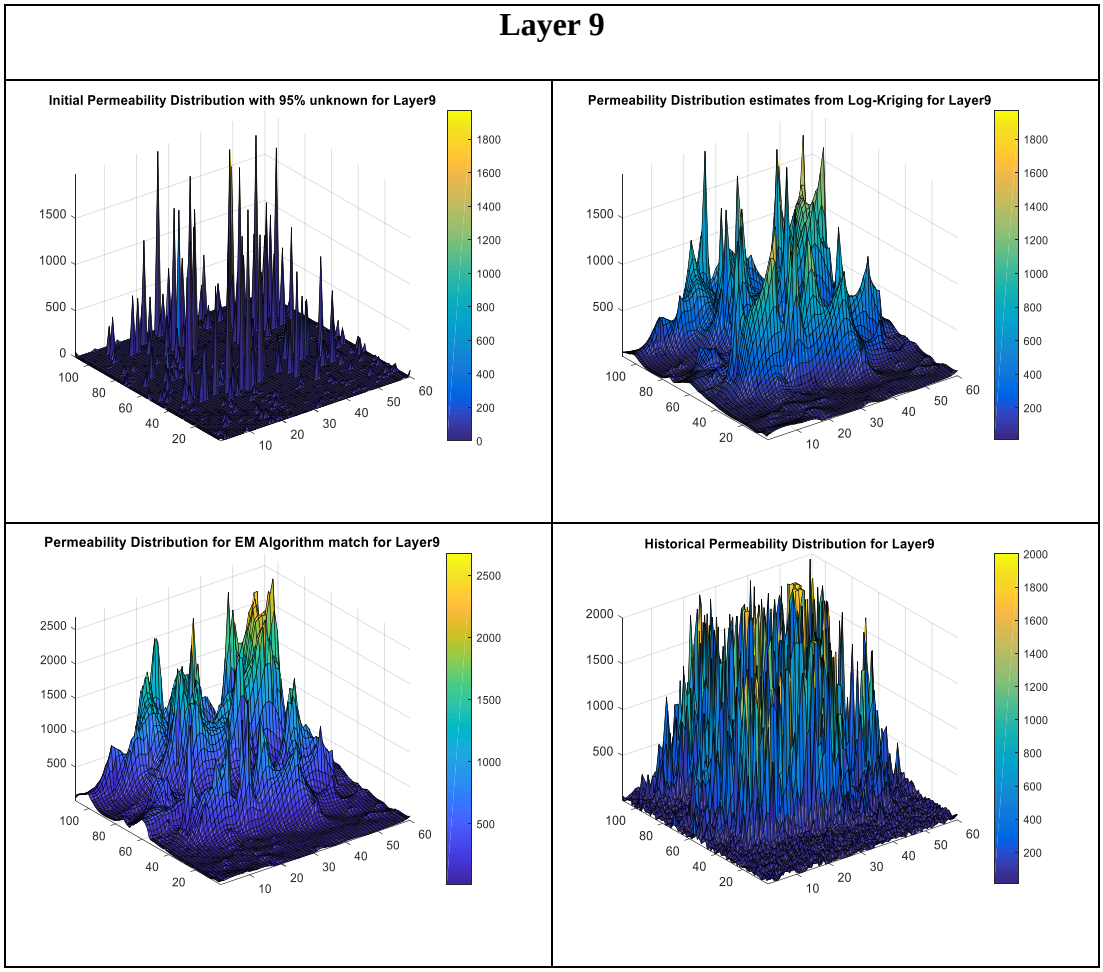


Figure 45: Permeability distribution at different workflow steps in comparison to the true permeability distribution Layer 9 of the SPE10 model

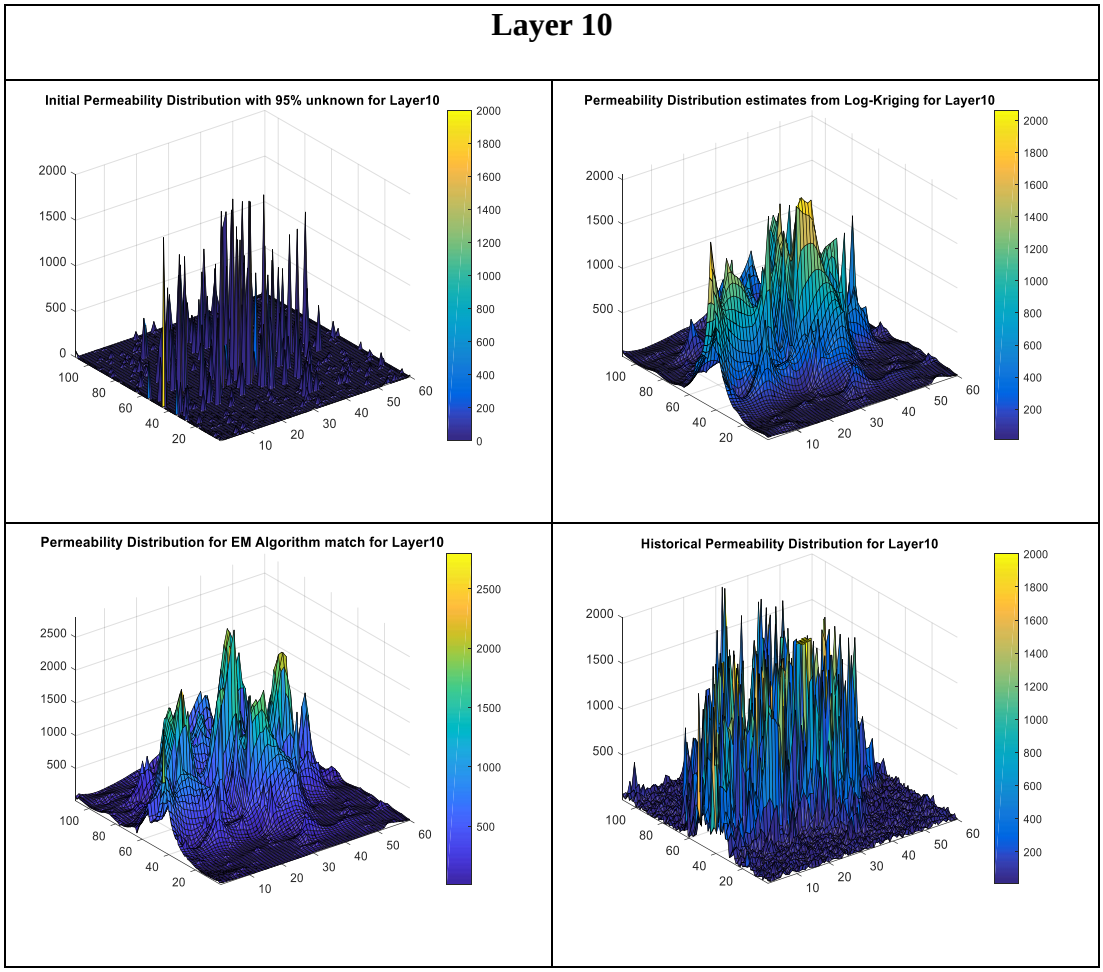


Figure 46: Permeability distribution at different workflow steps in comparison to the true permeability distribution Layer 10 of the SPE10 model

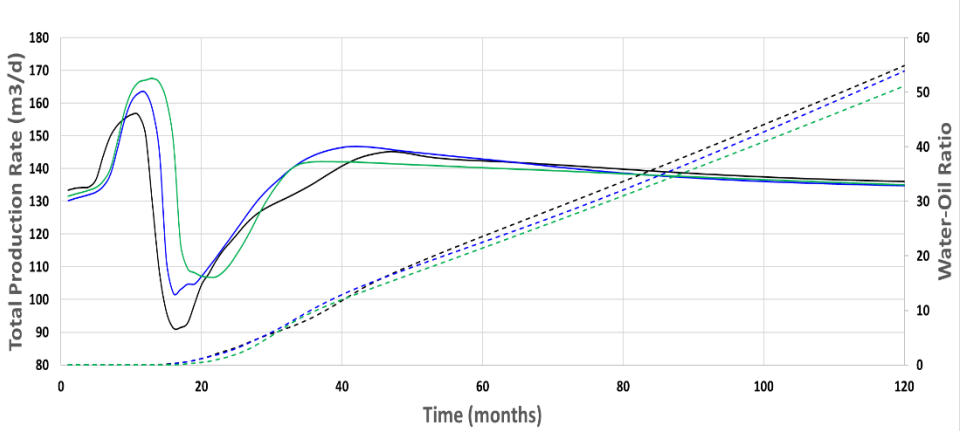
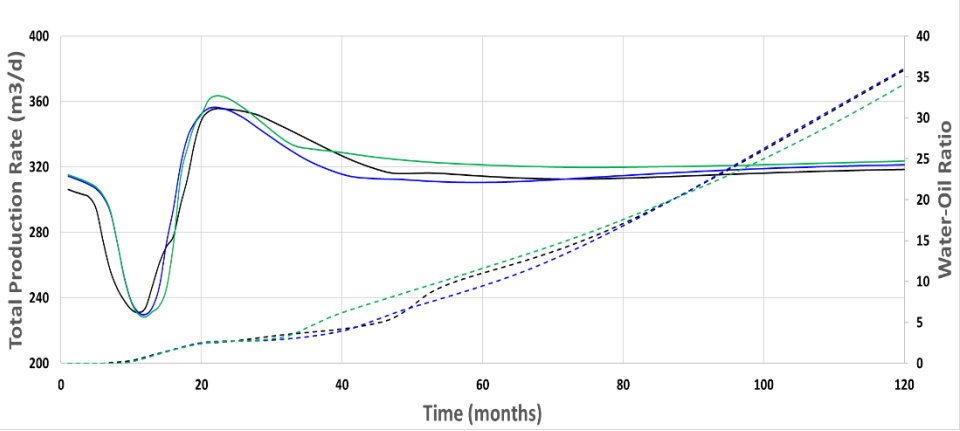
Non-Uniqueness of History Matching Method

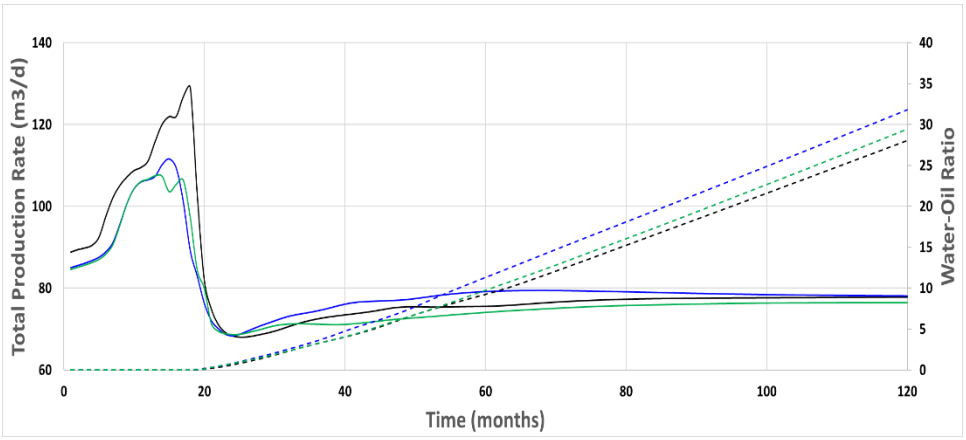
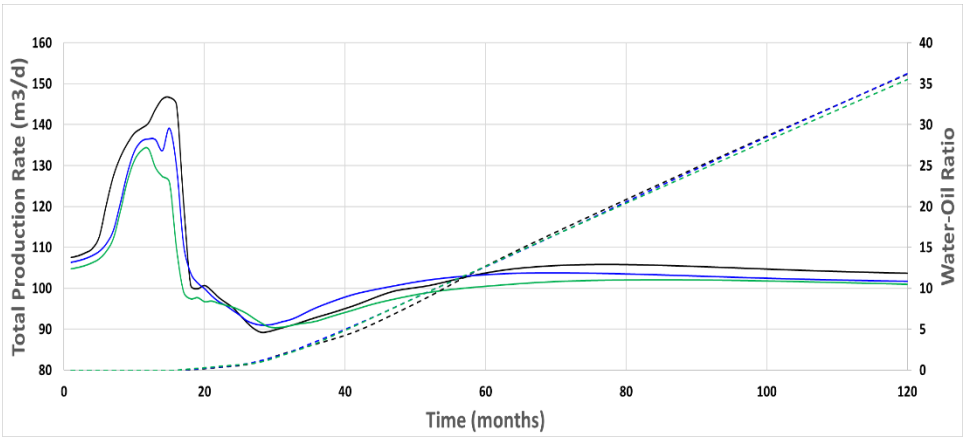
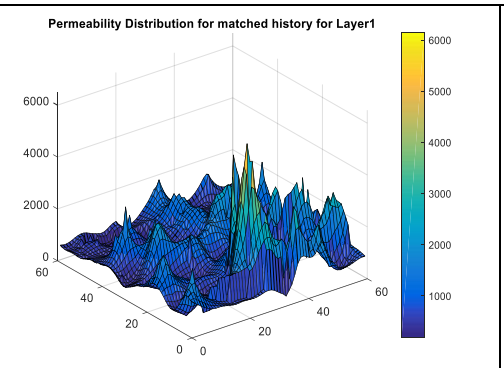
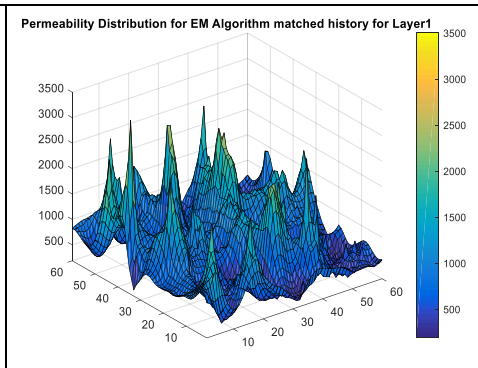
To further illustrate the non-uniqueness of both the Normalized Error method and the EM Algorithm method we do so with the Egg model and show that we get very good history matches with both methods as is shown in the Table 3 below. With the two different sets of permeability distributions, we were able to match the production history and the water-oil ratio for each of the producers. In Table 3 and Table 4, we have shown that for each producer, we get very good history match using the Normalized Error method and the EM Algorithm method for the Egg model and SPE10 model respectively. However, a review of the permeability distribution of each layer, shows that the permeability distribution of each layer for Normalized Error method and EM Algorithm method are very different.

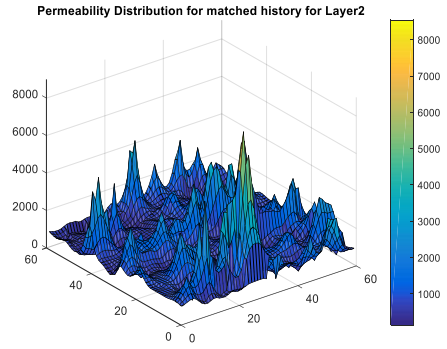
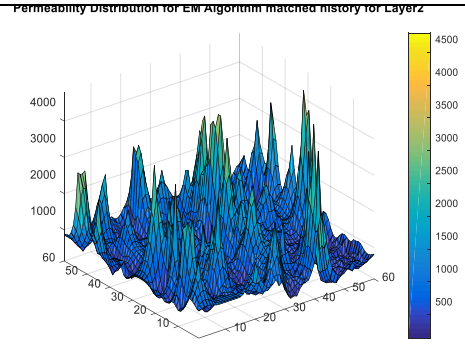
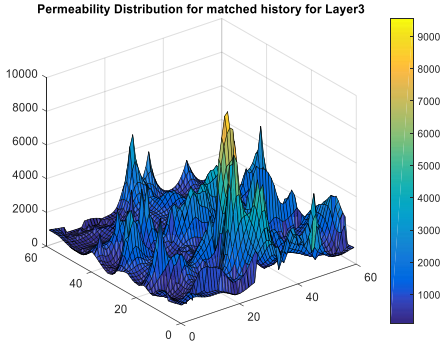
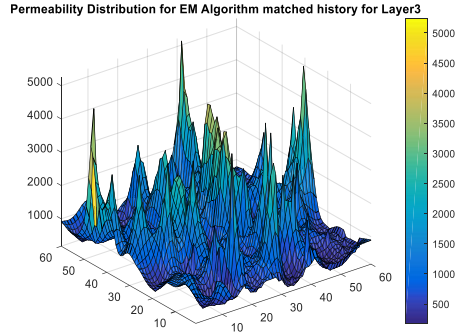
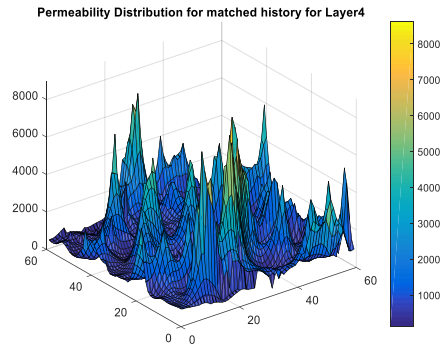
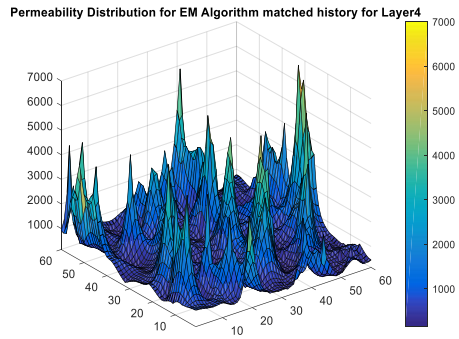
However, it is important to note that we were able to get a match in 10 loops within 0.4hrs, much quicker than with the EM Algorithm which took 20.5hrs in 5 loops. This is because, for each loop (outer loop) we go through, we have the following additional "inner" loops within the EM Algorithm:

1. Loop through each cell to get upper and lower bounds of permeability multipliers (This is the Normalized Error Workflow depict in Figure 16).
2. For each cell, once the permeability multipliers are calculated using Normalized error of the Water-Oil ratio and the mean error of the total production rate, loop to calculate and converge each to centroid in a cluster. Hence, the more data points (we used 100) per bound in a cell; the more clusters used, the longer the EM Algorithm will take to run even though it might converge to a match in less

number of “outer” loops. The EM Algorithm builds on the output of the Normalization Error method and is depicted in Figure 18.

	History Match with Normalized Error Method	**History Match with EM Algorithm Method**
# of Loops	10	5
Run Time	0.4 hrs.	20.5 hrs.
Prod1		
Prod2		

Prod3	
Prod4	
Layer1	 

Layer2		
Layer3		
Layer4		

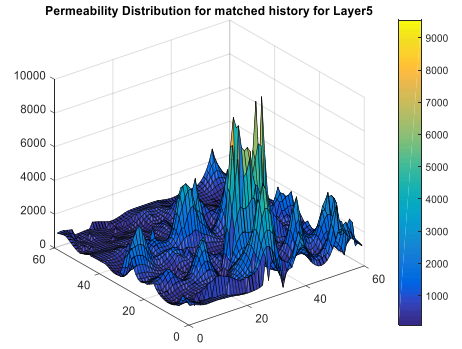
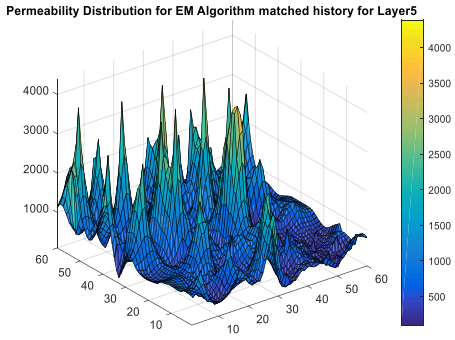
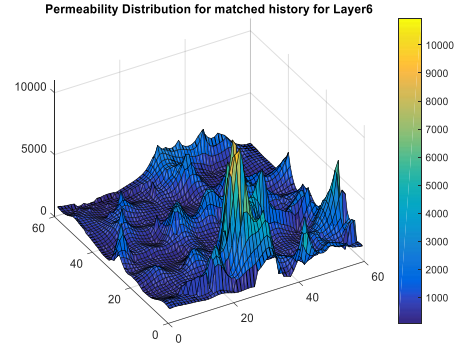
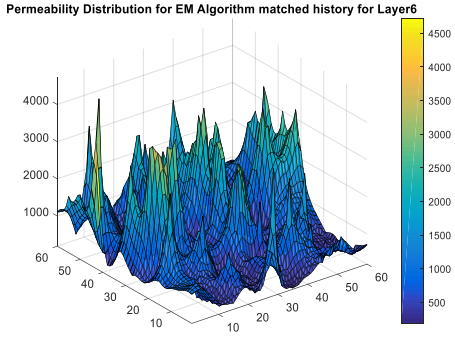
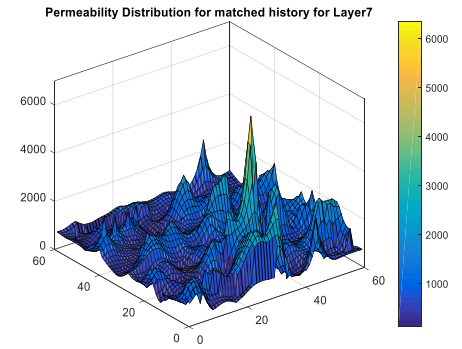
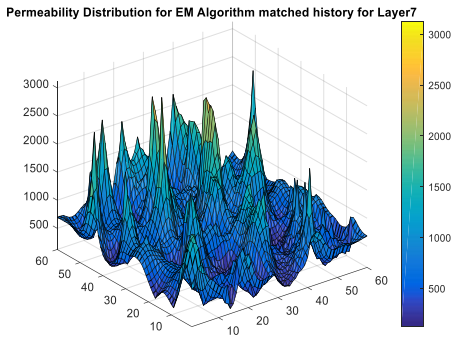
Layer5		
Layer6		
Layer7		

Table 3: Comparison of the Normalized Error Method and the EM Algorithm Method for History Matching of the Egg Model

Also, a comparison of the histograms of the field wide permeability distribution calculated via the EM Algorithm method as well as the Normalized Error method were compared to the true field wide permeability distribution. The following were observed:

- Both EM Algorithm and Normalized Error methods smoothen the variability in the model.
- The EM Algorithm method does less smoothening compared to the Normalized Error method
- Final result is dependent of our initial guess from log-Kriging

The reason why we experience more smoothening of the data with the Normalized Error method is that we only use the calculated maximum or minimum permeability multipliers (initial average permeability multipliers) to modify the unobserved contributing grid cell permeability. Whereas, with the EM Algorithm method, we search for the optimum permeability multiplier within a bound of the calculated maximum or minimum permeability multipliers and the previous permeability multiplier as depicted in Figure 18.

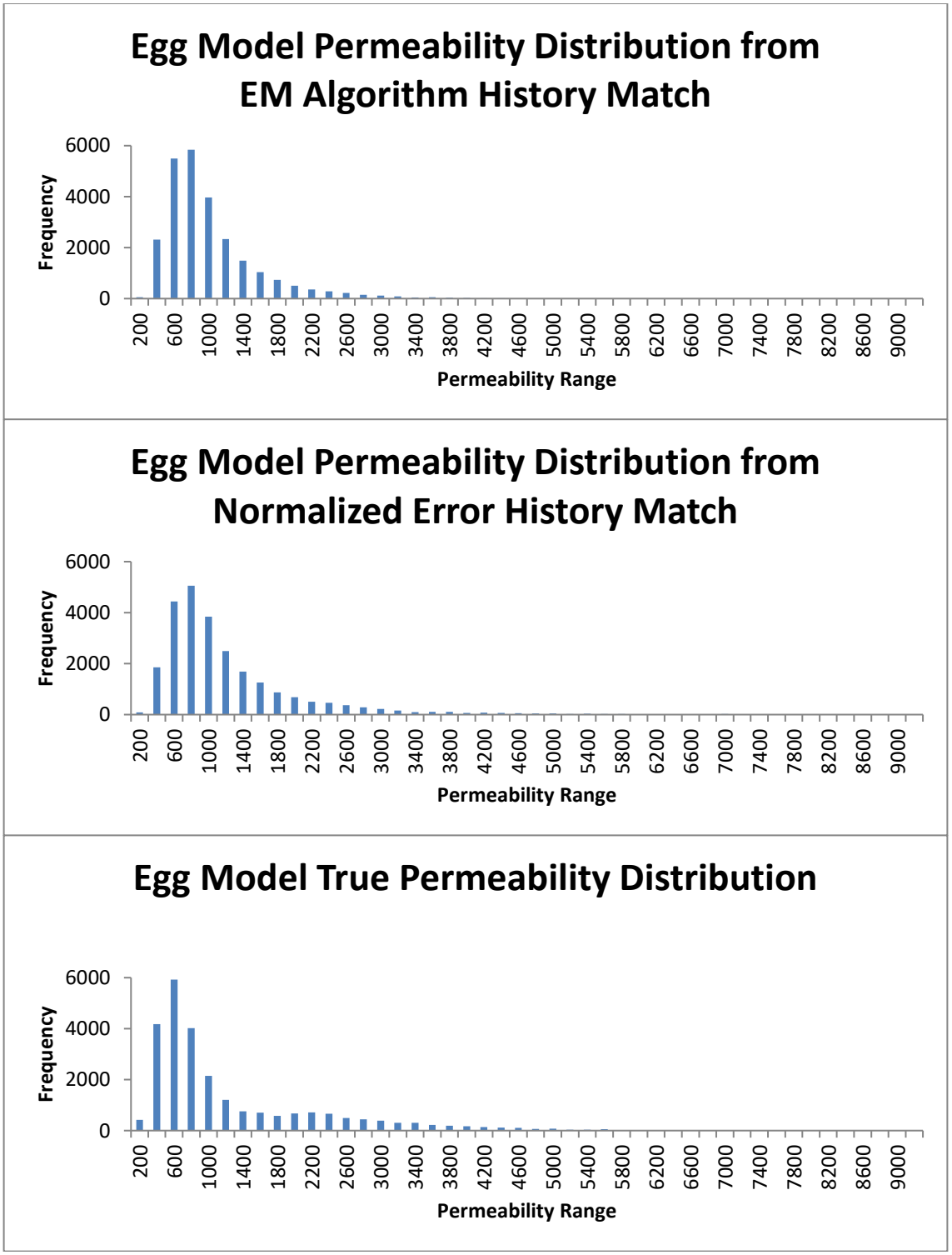
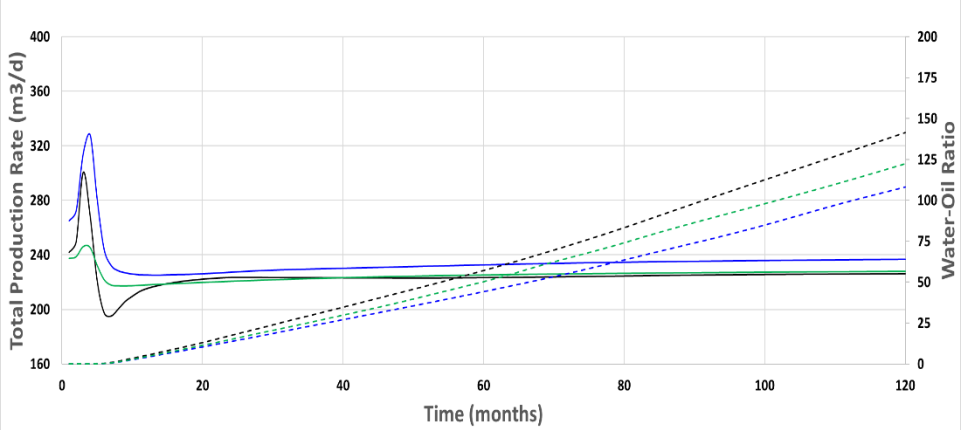
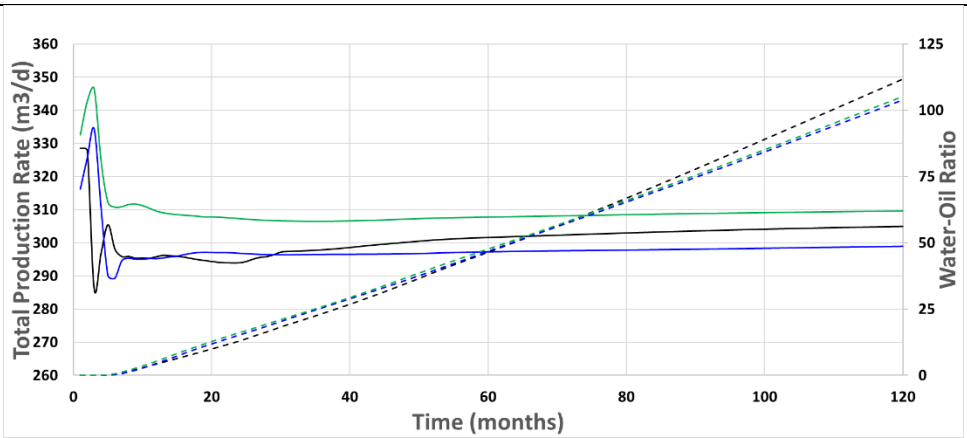
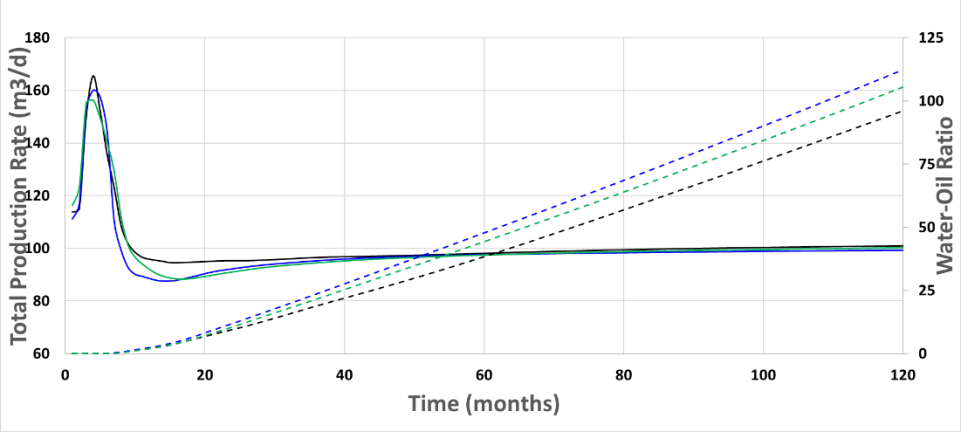
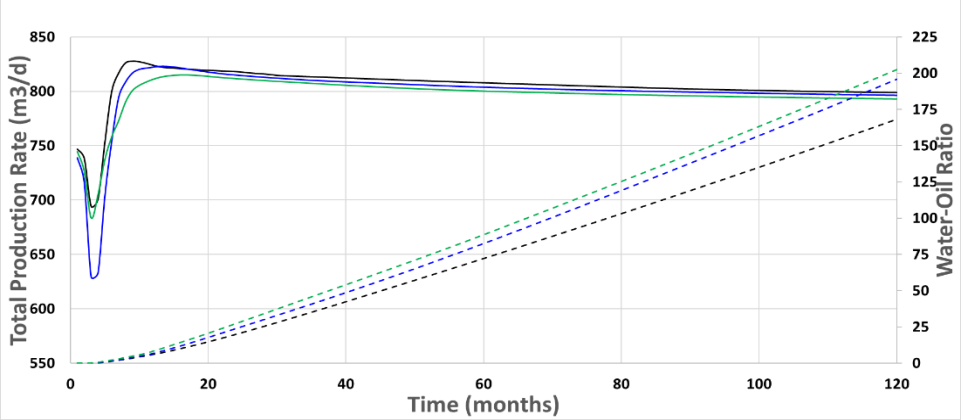
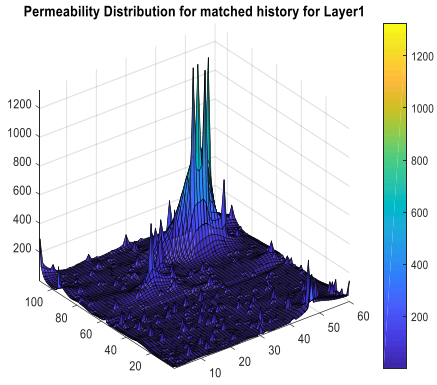
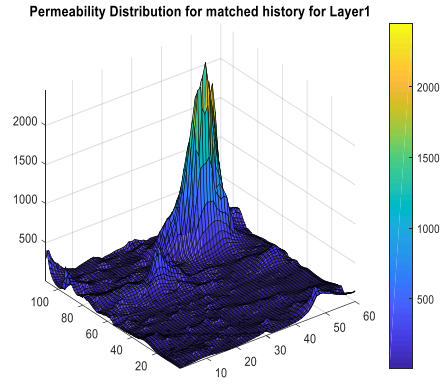
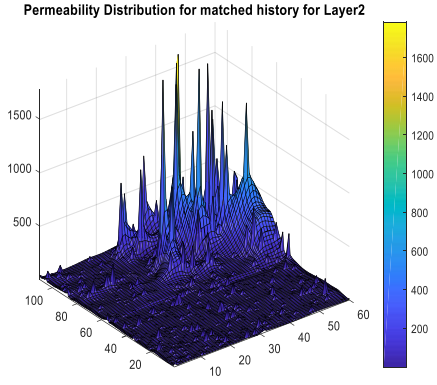
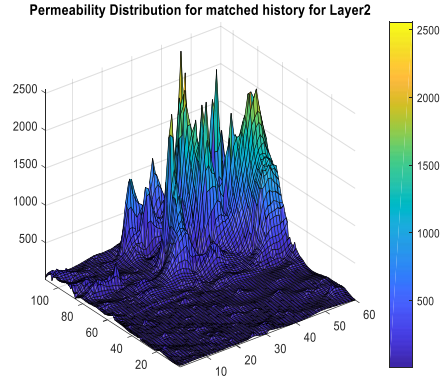
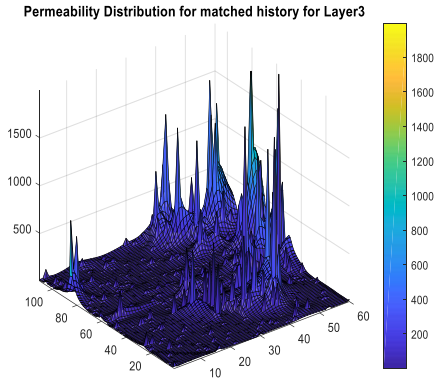
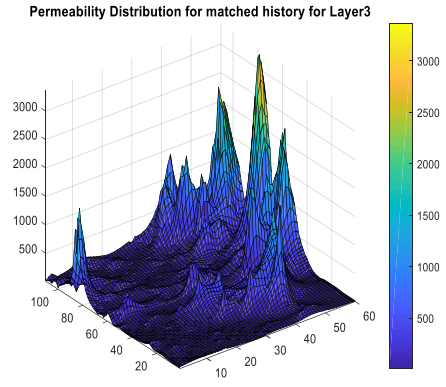


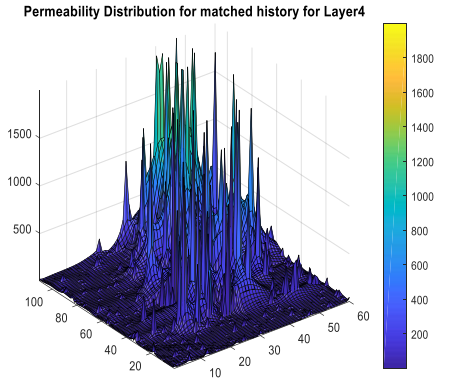
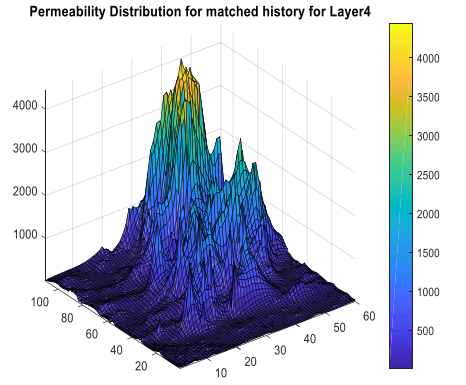
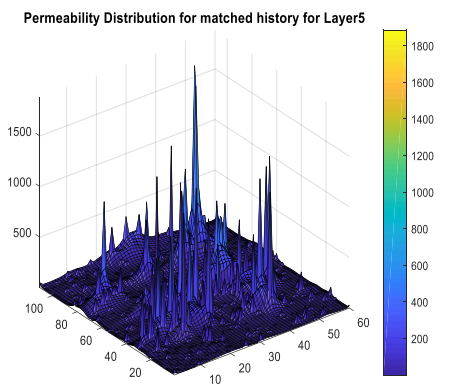
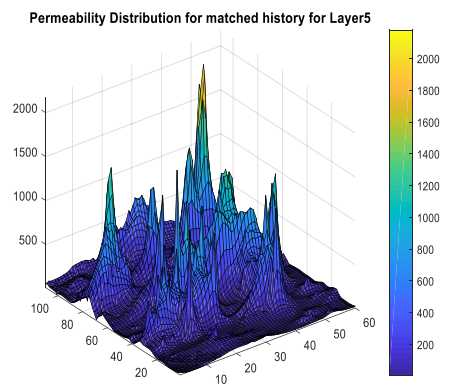
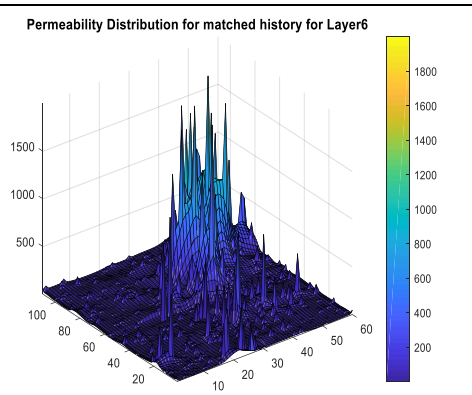
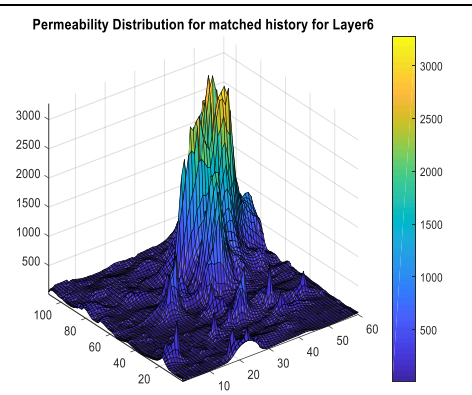
Figure 47: Comparison of Egg model Permeability Distribution from the EM Algorithm Method and Normalized Error Method with the true permeability distribution

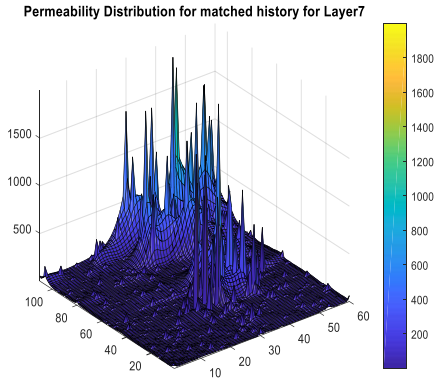
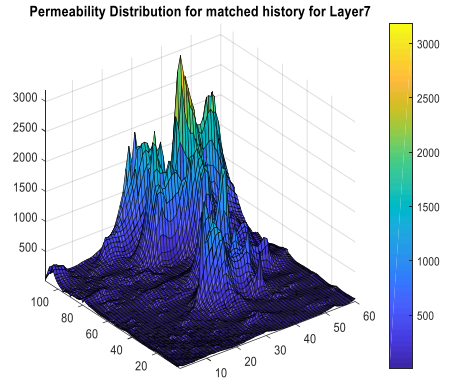
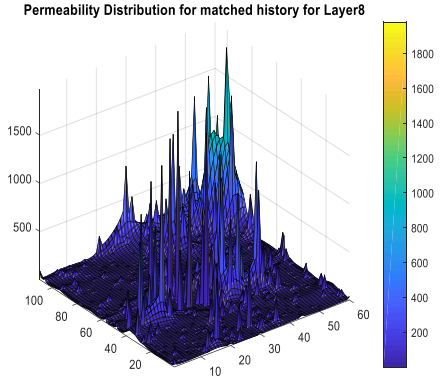
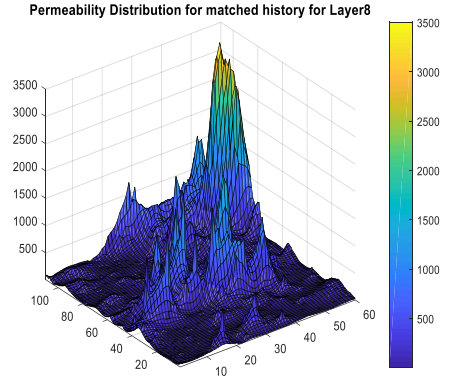
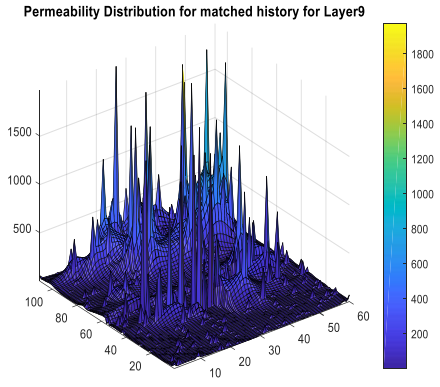
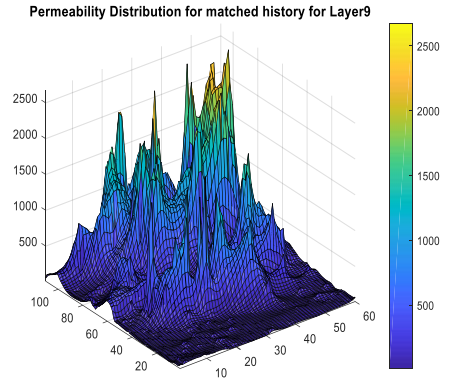
As shown with the Egg model, the non-uniqueness of both the Normalized Error method and the EM Algorithm method is further illustrated for the SPE10 model, and show that we get very good history matches with both methods as is shown in the tables below. With the two different sets of permeability distributions, we were able to also match the production history and the water-oil ratio for each of the producers.

	History Match with Normalized Error Method	**History Match with EM Algorithm Method**
# of Loops	52	10
Run Time	6.2 hrs.	166.7 hrs.
Prod1		

Prod2	
Prod3	
Prod4	

Layer1		
Layer2		
Layer3		

Layer4		
Layer5		
Layer6		

Layer7		
Layer8		
Layer9		

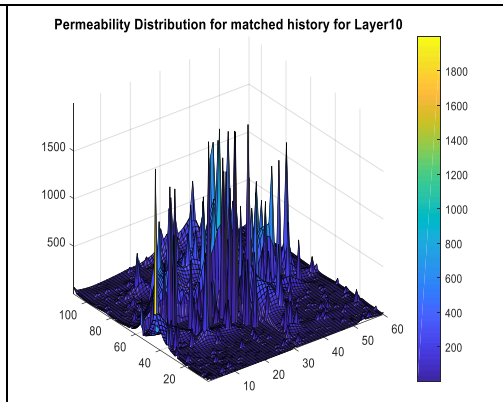
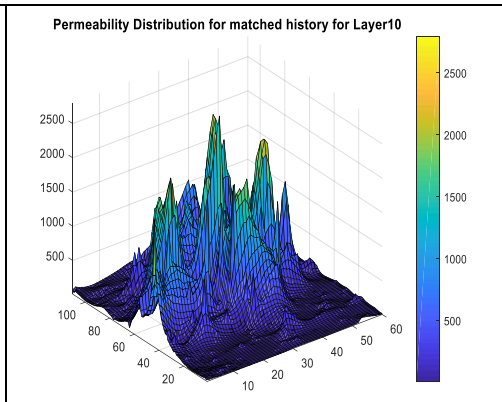
Layer10		

Table 4: Comparison of the Normalized Error Method and the EM Algorithm Method for History Matching of the SPE10 Model

Similar to the Egg model, a comparison of the histograms of the field wide permeability distribution calculated via the EM Algorithm method as well as the Normalized Error method were compared to the true field wide permeability distribution. The following similar observations were made:

- Both EM Algorithm and Normalized Error methods smoothen the variability in the model.
- The EM Algorithm method does less smoothening compared to the Normalized Error method.
- Final result is dependent of our initial guess from log-Kriging.

The reason why we experience more smoothening of the data with the Normalized Error method is as previously explained for the Egg model.

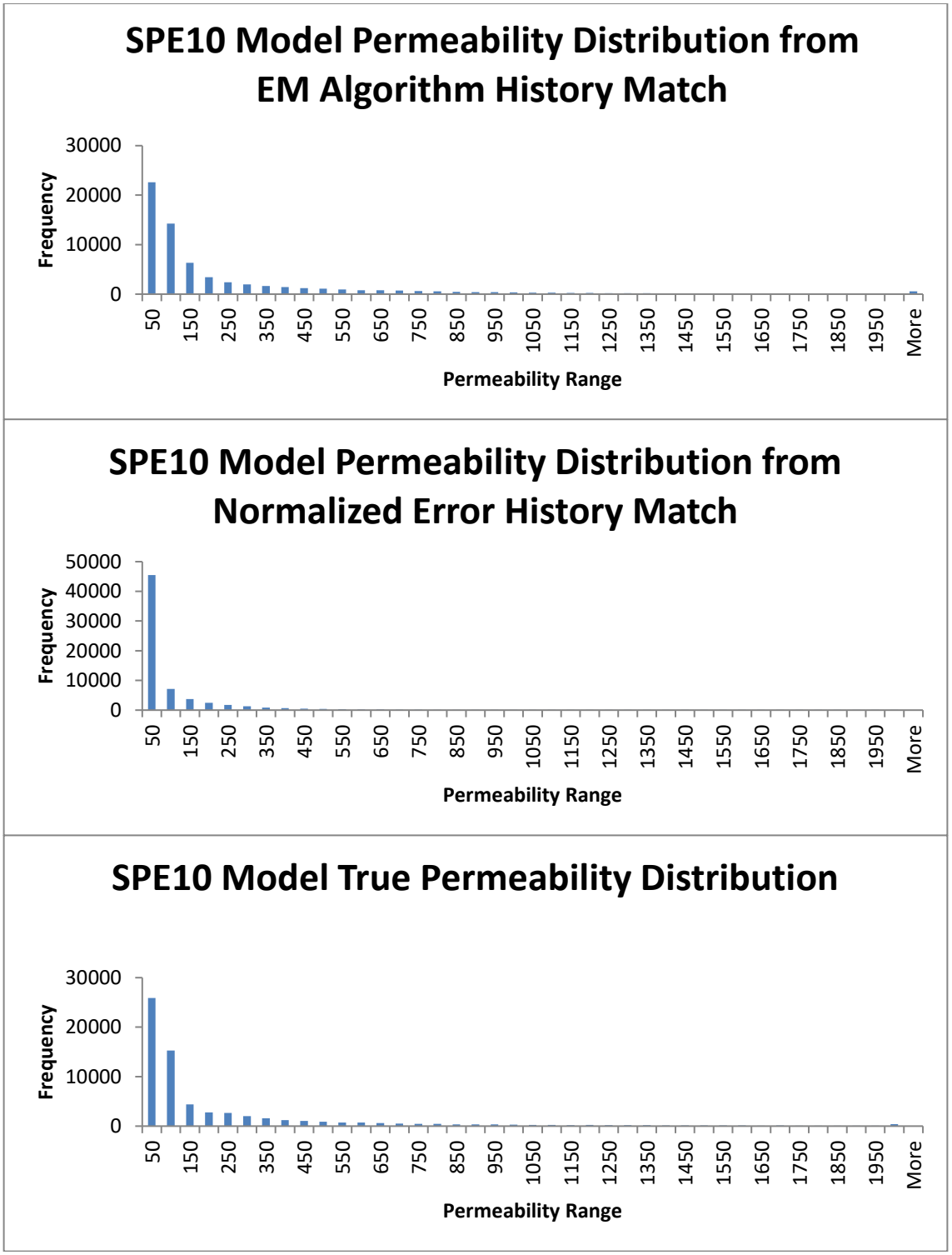


Figure 48: Comparison of SPE10 model Permeability Distribution from the EM Algorithm Method and Normalized Error Method with the true permeability distribution

Drawback of History Matching Method

Every history matching method has its own drawback. The methods described in this research are not exempted from that fact. A number of drawback of these history matching methods are hereby discussed.

1. A good match is a function of the quality of your initial permeability distribution estimation. The proximity of the initial permeability distribution estimates also affects the quality of the final permeability distribution match when compared to the true field-wide permeability distribution.
2. Method is not monotonically convergent: As seen in Figure 49 below, neither the Normalized method nor the EM Algorithm method are monotonically convergent. Although for the EM Algorithm, given that our permeability multipliers search is within bounded interval and consequently are sure to converge given that our search is always in a finite space within the bounds, the overall error of the history matching process may not converge below the threshold error limit simultaneously for all well. There seems to be no guaranty that a new error calculated will be less than the previous error. As a result, it is rather difficult to ascertain convergence with a good measure of certainty as depicted in Figure 49 below.

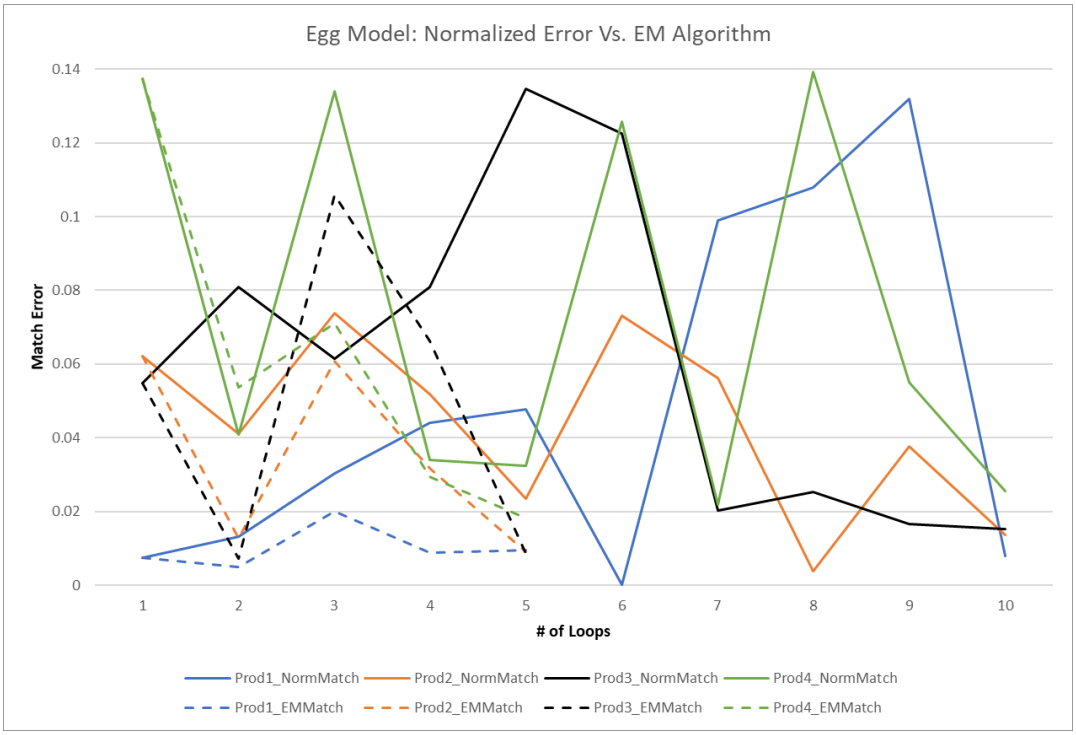


Figure 49: Uncertainty in the error reduction of both the Normalized Error and EM Algorithm methods

3. Given the inherent nature of the EM Algorithm in which we have multiple loops to converge to a permeability multiplier for a grid cell, this method can take time particularly when we have thousands of grid cells, and in real field scenarios, we may be looking at millions of grid cells to loop through, each grid cell having multiple loops of its own to calculate a permeability multiplier for it. The EM Algorithm method is computationally expensive and parallel computing solution or clever ways of reducing computational time needs to be thought-out.
4. Method searches for permeability multipliers that will simultaneously bring the errors of all the producers to a minimum. This approach may introduce a bit of a

challenge in matching the historical production given that if the error threshold is set for 2% and one of the computed production data is at 2.1% while others are below the 2% threshold, given that this approach is automatic in the very sense of it, we would not have met the history match criterion of 2% for all producers simultaneously and the algorithm would continue to search for a new multiplier that would make the errors of all the producers simultaneously below the error threshold as depicted by Figure 49.

Restrictions of History Matching Method

Given the nature of the history matching method, the following restrictions of the method are worth noting:

1. Particularly good for early development fields with less than 6 producers as explained by Figure 50. The ease with which the method can match historical production as a function of the number of producers whose history is to be match was investigated. It shows that the history matching methods have difficulty converging if there are more than 6 producers to match. Of course, this is a function of how homogeneous the permeability distribution is across the field. For this exercise, we have used the SPE10 model with a total of 9 producers. As noted, the SPE10 model is a highly heterogenous system and can be quite challenging to match. Having said that, the fact still remains that there will always be an increased difficulty in matching historical production within a

reasonable error threshold if the number of producers is increased, given that all errors must simultaneously be within the error threshold to have a successful match. The more producers we have, the more difficult it will be to simultaneously match historical production within a reasonable error limit. Figure 50 depicts our attempt to match history for up to 9 producers within error thresholds ranging from 2% to 10%. The graph shows the increasing difficulty to match history within error thresholds for increasing number of wells. It shows that we were unable to match history for 8 and 9 producers and barely for 7 producers for which the error threshold was increased to 10% for us to get a match. The figure shows that the tighter the error threshold, the more difficult it is to simultaneously match historical production.

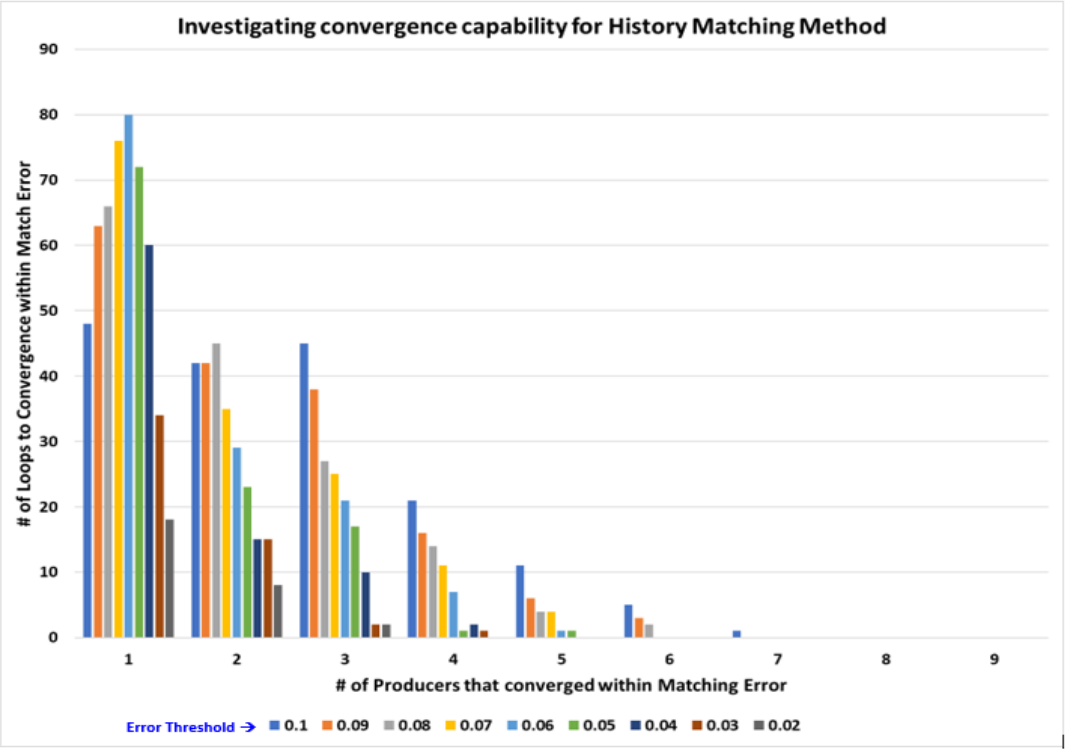


Figure 50: History Matching success as a function of number of wells

2. Depending on the model size, the EM Algorithm method may need significant computational resources and can benefit immensely from parallel computing. Using parallel computing, the optimum permeability multipliers within each permeability multiplier bound may be computed simultaneously using multiple clusters instead of computing the optimum permeability multipliers one contributing grid cell at a time.
3. The less heterogeneous the field the better the performance of this method. This is particularly true when the history matching approach we have undertaken is seen as a large optimization problem in which we perturb the permeability (using

the permeability multiplier) for each contributing grid cell in such a way as to minimize the total production error in all the producers simultaneously. This challenge of perturbing the permeability would be made much easier if the formation is less heterogeneous.

CHAPTER V

CONCLUSIONS

We have developed a method that successfully matches historical production in an oil-water system with sparse permeability data by:

- Utilizing the normalized water-oil ratio error and the mean production error between the calculated and historical production data to calculate the permeability multiplier bounds for each contributing grid cell, taking into consideration:
    a. The distance between each contributing grid cell and the producers; and
    b. The relative contribution of each grid cell to the producers, the process of which takes into consideration the geology of the earth model.
- Utilizing, for each contributing grid cell, the EM Algorithm to select the optimum permeability multiplier of each cluster of permeability multipliers within the permeability multiplier bounds.
- Calculating, for each contributing grid cell, the average of the resulting optimum cluster permeability multipliers as the final grid cell permeability multiplier, and calculating a new permeability of each contributing grid cell.
- Repeating the procedure until the error between the calculated and historical production data, for all producers, are simultaneously below a defined threshold.

Future Research Opportunities

The history matching methods developed in this research is in its infancy and will benefit from a wide range of future research opportunities. Some of which are noted below. Unlike many other methods like the streamline method, ensemble Kalman filter method, to mention a few, that have benefitted from many years of research, this method is new and hope that these methods will benefit from good quality research in the many years to come.

Some of the future opportunities for improvement include, but are not limited to, the following:

1. Increased computational speed:
    a. The EM algorithm section of this work can greatly benefit from parallel computing so that we can compute optimum permeability multipliers for multiple contributing grid cells at once instead of having to do compute one grid cell at a time.
    b. Instead of computing the permeability multiplier one grid cell at a time, the method can benefit from sectioning the entire field into different regions and possibly sub-regions such that we will only need to compute a permeability multiplier for the region or sub-region. This approach will mean a considerably less number of permeability multipliers will need to be computed, hence tremendously reducing the computational time.

2. For this work, we have only focused on oil-water system that is incompressible. There may be opportunity to investigate this history matching method in a three-fluid (oil-water-gas) system.
3. We didn't investigate the matching of field pressure during this research. However, the possibility of matching the historical bottom hole pressure and field pressure while simultaneously matching the historical production rate can be a subject of future investigation.
4. In the current research, we have explored the matching of historical production rate by making changes to the permeability distribution only by updating the permeability multiplier for each contributing grid cell. In reality, there will always be a need to make changes to more than one parameter to match historical production rate. Future work may be conducted to provide the ability to change multiple parameters for history matching.

REFERENCES

1. Nakayasu M., Goda T., Tanaka K., and Sato K. (2016). Evaluating the Value of Single-Point Data in Heterogeneous Reservoirs with the Expectation-Maximization Algorithm: Society of Petroleum Engineers, SPE 179721.
2. Schlanser K., Grana D., and Campbell-Stone E. (2014). Petro-elastic Facies Classification in the Marcellus Shale by Applying Expectation-Maximization to Measured Well-Logs: Society of Exploration Geologists, DOI http://dx.doi.org/10.1190/segam2014-0939.1.
3. Sassen D. and Lasscock B. (2015). A Pre-Stack Seismic Inversion with L1 Constraints and Uncertainty Estimation Using the Expectation Maximization Algorithm: Society of Exploration Geologists, DOI http://dx.doi.org/10.1190/segam2015-5912098.1.
4. Dempster A.P., Laird N.M., Rubin D.B. (1977). Maximum Likelihood from Incomplete Data via EM Algorithm: Journal of the Royal Statistical Society, Series B **39** (1): 1–38. JSTOR 2984875.
5. Dellaert F. (2002). The Expectation Maximization Algorithm: College of Computing, Georgia Institute of Technology, Technical Report number GIT-GVU-02-20.
6. Sundberg R. (1974). Maximum likelihood theory for incomplete data from an exponential family: Scandinavian Journal of Statistics 1 (2): 49–58. JSTOR 4615553.

7. Schon T.B. (2009). An Explanation of the Expectation Maximization Algorithm: Automatic Control at Linkopings Universitet, Report no: LiTH-ISY-R-2915.
8. Neal R.M. and Hinton G.E. (1999). A view of the EM algorithm that justifies incremental, sparse, and other variants: Learning in Graphical Models (Cambridge, MA: MIT Press): 355–368. ISBN 0-262-60032-3.
9. Einicke G.A., Falco G., and Malos J.T. (2010). EM Algorithm State Matrix Estimation for Navigation: IEEE Signal Processing Letters 17 (5): 437–440.
10. Hastie T., Tibshirani R., and Friedman, J. (2001). 8.5 The EM algorithm: The Elements of Statistical Learning. New York: Springer. pp. 236–243.
11. Zhu X. (2007). The EM Algorithm. The Department of Computer Science, University of Wisconsin – Madison.
12. Vaida F. (2005). Parameter Convergence for EM and MM Algorithms: Statistica Sinica 15 (2005), 831-840.
13. Xu L. and Jordan M. (1995). On Convergence Properties of the EM Algorithm for Gaussian Mixtures: MIT Artificial Intelligence Laboratory and Center for Biological and Computational Learning, A.I. Memo No. 1520, and C.B.C.L. Paper No. 111.
14. Wu, C. F. Jeff (1983). On the Convergence Properties of the EM Algorithm: Annals of Statistics 11 (1): 95–103.
15. Jamshidian M., Jennrich R.I. (1997). Acceleration of the EM Algorithm by using Quasi-Newton Methods. Journal of the Royal Statistical Society, Series B 59 (2): 569–587.
16. Yin J., Zhang Y., and Gao L. (2012). Accelerating Expectation-Maximization Algorithms with Frequent Updates: IEEE International Conference on Cluster Computing. DOI 10.1109/CLUSTER 2012.81.

17. Meng X., Rubin D.B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework: Biometrika 80 (2): 267–278.
18. Borman S. (2004). The Expectation Maximization Algorithm – A Short Tutorial
19. Roche A. (2003). EM Algorithm and Variants: An Informal Tutorial: arXiv:1105.1476v2 [stat.CO].
20. Bilmes J. (1998). A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models: International Computer Science Institute, TR-97-021.
21. Gupta M.R. and Chen Y. (2013). Theory and Use of the EM Algorithm (Foundations and Trends(r) in Signal Processing: ISBN-13: 978-1601984302.
22. McLachlan G. and Krishnan T. (2008). The EM Algorithm and Extensions. ISBN: 978-0-471-20170-0.
23. Ross S. (2003). Introduction to Probability Models: ISBN 0-12-598055-8.
24. Sergeev A.S. and Arapova R.K. (2002). The use of the Expectation-Maximization (EM) Algorithm for Maximum Likelihood Estimation of Gametic Frequencies of Multilocus Polymorphic Codominant Systems Based on Sampled Population Data: Genetika. 2002 Mar;38(3):407-18.
25. Hudson H.M. and Larkin R.S. (1994). Accelerated Image Reconstruction Using Ordered Subsets of Projection Data: IEEE Transactions on Medical Imaging, Vol. 13, No. 4, December 1994.

26. Forero P.A., Cano A. and Giannakis G.B. (2008). Consensus-Based Distributed Expectation-Maximization Algorithm for Density Estimation and Classification using Wireless Sensor Networks. IEEE 1-4244-1484-9/08.
27. Chen Y. and Gupta M.R. (2010). EM Demystified: An Expectation Maximization Tutorial. UWEE Technical Report Number UWEETR-2010-0002.
28. Andrieu C. and Doucet A. (2003). Online Expectation-Maximization Type Algorithm for Parameter Estimation in General State Space Models. IEEE Xplore. ISBN 0-7803-7663-3. DOI 10.1109/ICASSP.2003.1201620
29. Lipinski B., Herzog H., Kops E.R., Oberschelp W. and Muller-Gartner H.W. (1997). Expectation-Maximization Reconstruction of Positron Emission Tomography Images Using Anatomical Magnetic Resonance Information. IEEE S 0278-0062(97)02400-2.
30. Thrun S., Martin C., Liu Y., Hahnel D., Emery-Montemerlo R., Chakrabarti D. and Burgard W. (2004). A Real-Time Expectation Maximization Algorithm for Acquiring Multiplanar Maps of Indoor Environments with Mobile Robots. IEEE 1042-296X/04.
31. Frenkel L. and Feder M. (1999). Recursive Expectation-Maximization (EM) Algorithm for Time-Varying Parameters with Applications to Multiple Target Tracking. IEEE 1053-5875x/99.
32. Georghiades C.N. and Snyder D.L. (1991). The Expectation-Maximization Algorithm for Symbol Unsynchronized Sequence Detection. IEEE 0090-6778/91/0100-0054.

33. Bowsher J.E. and Floy C.E. Jr. (1990). Treatment of Compton Scattering in Maximum Likelihood, Expectation-Maximization Reconstruction of SPECT Images. Journal of Nuclear Medicine 1991; 32:1285-1291.

34. Cao N., Nehorai A. and Jacob M. (2007). Image Reconstruction for Diffuse Optical Tomography using Sparsity Regularization and Expectation-Maximization Algorithm. Oct 2007/Vol. 15, No. 21 / OPTICS EXPRESS 13695.

35. Cappe O. and Moulines E. (2009). On-line Expectation-Maximization Algorithm for Latent Data Models. Journal of the Royal Statistical Society B (2009) **71**, Part 3, pp. 593-613.

36. Tom B.C. and Katsaggelos A.K. (1995). Reconstruction of a High-Resolution Image by Simultaneous Registration, Restoration, and Interpolation of Low-Resolution Images. IEEE 0-8186-7310-9/95.

37. Marinakis D., Dudek G. and Fleet D.J. (2005). Learning Sensor Network Topology through Monte Carlo Expectation Maximization. IEEE 0-7803-8914-X/05.

38. Caffo B.S., Jank W. and Jones G.L. (2005). Ascent-based Monte Carlo Expectation Maximization. Journal of the Royal Statistical Society B (2005) **67**, Part 2, pp. 235-251.

39. Stanley B.J. and Gulochom G. (1993). Numerical Estimation of Adsorption Energy Distributions from Adsorption Isotherm Data with the Expectation-Maximization Method. Journal of Physical Chemistry **1993**, 97, 8098-8104.

40. Farag A.A., El-Baz A. and Gimel'farb G. (2004). Density Estimation Using Modified Expectation-Maximization Algorithm for a Linear Combination of Gaussians. IEEE 0-7803-8554-3/04.
41. Tom B.C., Lay K. and Katsaggelos A.K. (1996). Multichannel Image Identification and Restoration Using the Expectation-Maximization Algorithm. Optical Engineering/January 1996/Vol. 35 No. 1/241.
42. Borran M.J. andNasiri-Kenari M. (2000). An Efficient Detection Technique for Synchronous CDMA Communication Systems Based on the Expectation-Maximization Algorithm. IEEE 0018-9545/00.
43. Lagendijk R.L., Biemond J. and Boekee D.E. (1990). Identification and Restoration of Noisy Blurred Images Using the Expectation-Maximization Algorithm. IEEE 0096-3518/90/0700.
44. Graca J.V., Ganchev K. and Taskar B. (2007) Expectation Maximization and Posterior Constraints. University of Pennsylvania ScholarlyCommons 12-1-2007.
45. Rohlfing T., Russakoff D.B. and Maurer C.R. Jr. (2004). Performance-Based Classifier Combination in Atlas-Based Image Segmentation Using Expectation-Maximization Parameter Estimation. IEEE 0278-0062/04.
46. Carson C., Belongie S., Greenspan H. and Malik J. (2002). Blobworld: Image Segmentation Using Expectation-Maximization and It Application to Image Querying. IEEE 0162-8828/02.
47. Bailey T.L. and Elkan C. (1995). Unsupervised Learning of Multiple Motifs in Biopolymers Using Expectation Maximization. Machine Learning, 21, 51-80 (1995).

48. Excoffier L. and Slatkin M. (1995). Maximum-Likelihood Estimation of Molecular Haplotype Frequencies in a Diploid Population. Mol. Biol. Evol. 12(5):921-927, 1995.

49. Zhang Y., Brady M. and Smith S. (2001). Segmentation of Brain MR Images Through a Hidden Markov Random Field Model and the Expectation Maximization Algorithm. IEEE 0278-0062/01.

50. Ranganath M.V., Dhawan A.P. and Mulani N. (1988). A Multigrid Expectation Maximization Reconstruction Algorithm for Positron Emission Tomography. IEEE 0278-0062/88/1200-0273.

51. Levitan E. and Herman G.T. (1987). A Maximum A Posteriori Probability Expectation Maximization Algorithm for Image Reconstruction in Emission Tomology. IEEE 0278-0062/87/0900-0185.

52. Lawrence C.E. and Reilly A.A. (1990). An Expectation Maximization (EM) Algorithm for the Identification and Characterization of Common Sites in Unaligned Biopolymer Sequences. PROTEINS: Structure, Function, and Genetics 7:41-51 (1990).

53. Ceppellini R., Siniscalco M. and Smith C.A.B. (1955). The Estimation of Gene Frequencies in a Random-Mating Population. Annals of Human Genetics Vol. 20 Issue 2, Nov. 1955 Pages 97-115.

54. Baum L.E. Petrie T., Soules G. and Wiess N. (1970). A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. The Annals of Mathematical Statistics. 1970, Vol. 41 No. 1, 164-171.

55. Zhang J. and Liang F. (2010). Robust Clustering Using Exponential Power Mixtures. BIOMETRICS 66, 1078-1086, December 2010.

56. Bondell H.D., Krishna A. and Ghosh S.K. (2010). Joint Variable Selection for Fixed and Random Effects in Linear Mixed-Effects Models. BIOMETRICS 66, 1069-1077, December 2010.

57. Shahkarami A., Mohaghegh S.D. and Hajizadeh Y. (2015). Assisted History Matching Using Pattern Recognition Technology. Society of Petroleum Engineers SPE-173405.

58. Shahkarami A., Mohaghegh S.D., Gholami V. and Haghighat S.A. (2014). Artificial Intelligence (AI) Assisted History Matching. Society of Petroleum Engineers SPE-169507.

59. Hajizadeh Y. (2010). Ants Can Do History Matching. Society of Petroleum Engineers SPE-141137

60. Dadashpour M. (2009). Reservoir Characterization Using Production Data and Time-Lapse Seismic Data. PhD Dissertation, The Department of Petroleum Engineering and Applied Geophysics, Norwegian University of Science and Technology (NTNU).

61. Hajizadeh Y., Nghiem L., Mirzabozorg A., Yang C., Li H. and Sousa M.C. (2013). A Soft and Law-Abiding Framework for History Matching and Optimization under Uncertainty. Society of Petroleum Engineers SPE-163636.

62. Aissaoui K. and Moreno M. (2013). A Comprehensive Approach for History Matching A Giant and Complex Oil Reservoir. Society of Petroleum Engineers SPE-165990.

63. Yang P.H. and Watson A.T. (1988). Automatic History Matching With Variable-Metric Methods. Society of Petroleum Engineers SPE-16977.

64. Cancelliere M., Verga F. and Viberti D. (2011). Benefits and Limitations of Assisted History Matching. Society of Petroleum Engineers SPE-146278.

65. Gang T. and Kelkar M. (2006). Efficient History Matching in Naturally Fractured Reservoirs. Society of Petroleum Engineers SPE-99578.

66. Tavassoli Z., Carter J.N. and King P.R. (2004). Errors in History Matching. Society of Petroleum Engineers SPE-86883.

67. Kabir C.S., Chien M.C.H. and Landa J.L. (2003). Experiences With Automated History Matching. Society of Petroleum Engineers SPE-79670.

68. Sayyafzadeh M. (2015). History Matching by Online Metamodeling. Society of Petroleum Engineers SPE-175618.

69. Singh A., Maucec M. and Knabe S. (2014). History Matching Using Streamline Trajectories. Society of Petroleum Engineers SPE-172146

70. Almeida Netto S.L., Schiozer D.J., Ligero E.L. and Maschio C. (2003). History Matching Using Uncertainty Analysis. Petroleum Society Canadian Institute of Mining, Metallurgy and Petroleum. Paper 2003-145.

71. Schiozer D.J., Almeida Netto S.L., Ligero E.L. and Maschio C. (2005). Integration of History Matching and Uncertainty Analysis. JCPT. July 2005, Vol. 44, No. 7.

72. Watkins A.J. (1993). Linking Universal Kriging and Reservoir History Matching. Society of Petroleum Engineers SPE-26610.

73. Abdel-Rahman M.R., Dharmawan M.A., Andelrahim R., Akhtar M.N., Mekki N., Al-Namlah S.A. (2013). New Approach to Validate History Matching Process. Society of Petroleum Engineers SPE-167654.

74. Yamada T. (2000). Non-Uniqueness of History Matching. Society of Petroleum Engineers SPE-59434.

75. Macmillan D.J., Pletcher J.L. and Bourgeois S.A. (1999). Practical Tools To Assist History Matching. Society of Petroleum Engineers SPE-51888.

76. Gavalas G.R., Shah P.C. and Seinfeld J.H. (1976). Reservoir History Matching by Bayesian Estimation. Society of Petroleum Engineers SPE-5740.

77. Tzu-hao Y., Jimenez E.A., Van Essen G., Chen C., Jin L, Girardi A., Gelderblom P., Horesh L. and Conn A.R. (2014). Reservoir Uncertainty Quantification Using Probabilistic History Matching Workflow. Society of Petroleum Engineers SPE-170893.

78. Couvreur C. (1997) The EM Algorithm: A Guided Tour. Computer Intensive Methods in Control and Signal Processing. PP 209-222. ISBN 978-1-4612-7373-8.

79. Hartley H.O. (1958) Maximum Likelihood Estimation from Incomplete Data. Biometrics, Vol. 14, No. 2 (Jun., 1958), pp. 174-194.

80. Shahvali M., Mallison B., Wei K, and Gross H. (2012) An Alternative to Streamlines for Flow Diagnostics on Structured and Unstructured Grids. Society of Petroleum Engineers SPE-146446.

81. Cheng H. (2005) Fast History Matching of Finite-Difference Model, Compressible and Three-Phase Flow using Streamline-Derived Sensitivities. Dissertation, Texas A&M University

82. Chuong B. and Serafim B. (2008). What is the expectation maximization algorithm? Nature Biotechnology 26, 897–899.

83. http://vis.supstat.com/2012/11/contour-plots-with-d3-and-r/

84. Bansal Y., Ertekin T., Karpyn Z., Ayala L., Suleen, FNU, Balogun O., Liebmann D. and Qiang S. (2013). Forecasting Well Performance in a Discontinuous Tight Oil Reservoir Using Artificial Neural Networks. Society of Petroleum Engineers SPE-164542-MS.

85. Jansen J.D., Fonseca R.M., Kahrobaei S., Siraj M.M., Van Essen G.M. and Van den Hof P.M.J. (2014). The egg model – A geological ensemble for reservoir simulation. TUDelft. UUID:a5929b0f-323a-4d42-8042-cf355b6190cb. DOI:10.1002/gdj3.21.

86. Christie M., and Blunt M. (2001), SPE Comparative Solution Project. http://www.spe.org/web/csp/.

87. Model Calibration and Efficient Reservoir Imaging (MCERI) Industrial Research Consortium, Texas A&M University, http://www.pe.tamu.edu/mceri/overview.html.

88. Enerdeq. IHS Markit, https://my.ihs.com/energy.

89. Mongwe W., www.wilsonmongwe.co.za/undersanding-the-em-algorithm. Example of a pdf of a Gaussian Mixture.

90. Kelkar M. and Perez G. Applied Geostatistics for Reservoir Characterization. Copyright 2002 ISBN 1-55563-095-2.

91. Chen J., Tan X. and Zhang R. Inference for Normal Mixtures in Mean and Variance. Statistica Sinica Vol. 18, No. 2, April 2008 (pp. 443-465).

92. Walker G.J., Chen Z.J, and Shaker W. (2013). Combining a Surveillance Testing Workflow into the Assisted History Matching Process to Reduce Uncertainties in a SAGD Reservoir. Society of Petroleum Engineers SPE-165546-MS.

93. Stephen K.D and Macbeth C.E. (2006). Reducing Reservoir Prediction Uncertainty Using Seismic History Matching. Society of Petroleum Engineers SPE-100295-MS.

94. Denney, D. (2012). Improved History Matching and Uncertainty Characterization. JPT Senior Technology Editor 0412-0142-JPT.

95. Ertekin T., Abou-Kassem J.H. and King G.R. Basic Applied Reservoir Simulation. SPE Textbook Series Vol. 7. ISBN 978-1-55563-089-8.

96. Aziz K. and Settari A. Petroleum Reservoir Simulation. ISBN 978-0853347873.

97. Jovanovic B.S., Suli E. (2013). Analysis of Finite Difference Schemes: For Linear Partial Differential Equations with Generalized Solutions. Springer Series in Computational Mathematics 46. ISBN 978-1-44715-460-0.

98. Smith G.D. (1985). Numerical Solution of Partial Differential Equation: Finite Difference Methods. Oxford University Press. ISBN 978-0-19859-650-9.

99. Johnson C. (2009). Numerical Solution of Partial Differential Equations by the Finite Element Method. Dover Publications. ISBN 978-0-48646-900-3.

100. Mazumder S. (2016). Numerical Methods for Partial Differential Equations: Finite Difference and Finite Volume Methods. Elsevier Science. ISBN 978-0-12849-894-1.

101. Leem J., Lee K., Kang., Park Y., and Park J. (2015). History Matching with Ensemble Kalman Filter Using Fast Marching Method in Shale Gas Reservoir. Society of Petroleum Engineers SPE-176164-MS.

102. Gavala G.R., Shah P.C. and Seinfeld J.H. (1976). Reservoir History Matching by Bayesian Estimation. Society of Petroleum Engineers 5740-PA SPE Journal Paper.

103. Shams M., Al-Banbi A. and Sayyouh H. (2017). A Comparative Study of Experimental Design Techniques in Assisted History Matching. Society of Petroleum Engineers SPE-183679-MS.

104. Datta-Gupta A. and King M.J. (2007). Streamline Simulation: Theory and Practice. SPE Textbook Series Vol. 11. ISBN 978-1-55563-111-6.

105. Oliver D.S., Reynolds A.C. and Liu Ning. (2008). Inverse Theory for Petroleum Reservoir Characterization and History Matching. ISBN 978-0521881517.